Collective is Different: Information Exchange and Speed-Accuracy Trade-Offs in Self-Organized Patterning

Ashutosh Tripathi [1,2], Jörn Dunkel [1], and Dominic J. Skinner [2,*]

[1] *Department of Mathematics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA*

[2] *Center for Computational Biology, Flatiron Institute, 162 5th Avenue, New York, New York 10010, USA*

(Received 1 September 2025; accepted 5 January 2026; published 4 February 2026)

During development, highly ordered structures emerge as cells collectively coordinate with each other. While recent advances have clarified how individual cells process and respond to external signals, understanding collective cellular decision making remains a major challenge. Here, we introduce a minimal, analytically tractable model of cell patterning via local cell-cell communication. Using this framework, we identify a trade-off between the speed and accuracy of collective pattern formation and, by adapting techniques from stochastic chemical kinetics, quantify how information flows between cells during patterning. Our analysis reveals counterintuitive features of collective patterning: globally optimized solutions do not necessarily maximize intercellular information transfer and individual cells may appear suboptimal in isolation. Moreover, the model predicts that instantaneous information shared between cells can be nonmonotonic in time as patterning occurs. An analysis of recent experimental data from lateral inhibition in *Drosophila* pupal abdomen finds a qualitatively similar effect.

DOI: 10.1103/lfnt-8qbt

I. INTRODUCTION

Creating a functional organism requires complex multicellular coordination [1,2]. To this end, eukaryotic cells have evolved a multitude of mechanisms to sense external cues [1–3] and respond to them by changing their internal state [4–6], migrating [7], or signaling [2,8] to other cells. The underlying principles of multicellular self-organization remain elusive, despite numerous examples of developmental self-organization observed across biology [9–11]. Does development optimize for particular objectives, such as robustness or speed? If so, what constraints shape the outcomes? In the case of a single cell processing an exogenous signal, theoretical progress has been made to determine how individual cells could optimally process a static [4] or dynamic [12,13] signal, and how cells can sense their environment and optimally act upon it [14,15]. Many developmental contexts fit this paradigm: one population emits a signal and another responds without providing feedback to the sender. For example, in *Caenorhabditis elegans* vulval development an anchor cell secretes an EGF-like ligand which is received by vulval precursor cells but no reciprocal signal is sent [16,17]. Similarly, during ascidian development, vegetal cells secrete an FGF-like ligand, which induces a neural fate in specific animal cells and no reciprocal signal is sent [18,19].

Frequently, however, cells both send and receive signals, leading to complex nonlinear feedback [1,2,20]. Indeed, although the receiving cells in the previous examples do not feed back on the signal sender, they coordinate among themselves, through Notch signaling in *C. elegans* vulva [16] and EphrinA signaling in ascidian neural induction [21]. A single cell making noisy measurements of its external environment and acting on them can be modeled as a partially observed Markov decision process, for which optimal control is governed by the Bellman equation [15,22]. In contrast, interacting cells engaged in a collective task, such as patterning, can be represented as a decentralized partially observed Markov decision process, where computing an optimal strategy is generically hard, though finding good solutions is becoming increasingly practical [23,24]. Difficulties arising from decentralization are well known in computer science; for instance, the two generals' problem demonstrates that perfect coordination between decentralized actors is impossible when they communicate over noisy channels [25,26]. Echoing "more is different" [27,28], while multicellular systems are indeed made up of individual cells, the emergent principles that govern the self-organizing collective are not simple extensions of single-cell behavior. Development provides a proof by example that decentralized systems are capable of robust self-organization. Yet, despite numerous models demonstrating collective self-organization [10,29,30], the core principles, such as optimality and information flow, remain largely unexplored.

To study these fundamental questions on decentralized self-organization in an analytically and numerically tractable

*Contact author: dskinner@flatironinstitute.org

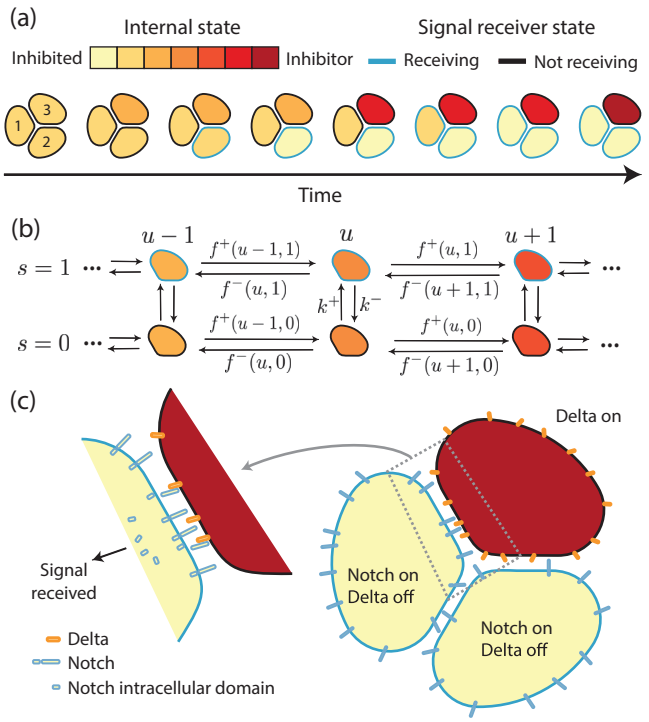

FIG. 1. Self-organization through lateral inhibition. (a) In our model, cells have an internal state abstractly ranging from an inhibitor state to an inhibited state, along with a receiver state which can be activated by neighboring cells. This allows cells to self-organize into a target pattern; here three cells organize so that there is one inhibitor cell and two inhibited cells. Throughout, we number cells counterclockwise starting from the leftmost cell; here cell 3 is the inhibitor. (b) Possible transitions for a single cell in the model and corresponding transition rates. Only the rate of receiving the signal, k^+, is a function of the neighboring cell states. (c) Delta-Notch signaling as a motivating biological example of lateral inhibition. When Delta (orange) on one cell binds to Notch (blue) on a neighboring cell, Notch is cleaved, releasing its intracellular domain and activating downstream transcriptional responses.

framework, we focus on a minimal model of self-organization through local cell-cell communication (Fig. 1). In our model, motivated by lateral inhibition [8,10,31], cells communicate imperfectly with their neighbors and are tasked with forming a pattern in which exactly one cell adopts an "inhibitor" state while its neighbors adopt an "inhibited" state. With this simple task we find that the patterning strategies obey a speed-accuracy trade-off: arbitrarily high patterning accuracy is possible, but at the cost of increasing the time required to pattern. By reframing the problem as a stochastic reaction network [32], we can precisely compute information-theoretic quantities, such as the mutual information shared between cells, allowing us to track how information flows in our model system. We contrast this dynamic information between trajectories with simpler instantaneous information quantities that can be computed from data, finding that instantaneous quantities can sometimes decrease as patterning occurs. Next, we show that after optimizing for a multicellular objective, individual cells do not appear to behave optimally when considered in isolation, and that the collectively optimal strategies neither maximize nor minimize the information transferred between cells. Finally, we connect these theoretical results to recent live-imaging experiments of collective patterning through Delta-Notch signaling [9], by computing the instantaneous mutual information from experimental measurements. We find, similarly to the model, that the instantaneous measures are nonmonotonic as patterning occurs.

II. COLLECTIVE PATTERNING MODEL

To explore the principles of communication and self-organization in a multicellular system, we take a model that accounts for the following essential biological features: (i) cells have some internal state, which could be used to classify cells into cell types or to specify a target pattern; (ii) cells have a way to communicate with each other; (iii) cells control their internal state based on signals they have received, and control the signals they send based on their internal state; and (iv) both sending and receiving of signals, as well as control of their internal state, are imperfect and subject to stochastic fluctuations. These criteria are general enough to describe a host of complex developmental feats of self-organization, from neural induction to digit specification. While development regularly features multiple signaling pathways, long-range morphogen diffusion, cell divisions, and physical rearrangements, for the sake of tractability we focus on the simpler example of a small fixed number of cells patterning through lateral inhibition with communication across physical cell-cell contacts.

A. Lateral inhibition as motivating example of collective cellular patterning

Lateral inhibition is a way to sort a population of cells into two states in a controlled manner by having cells in one state, or advancing toward that state, inhibit their neighbors from similarly advancing toward that state (Fig. 1). In development, this is often implemented through Notch signaling, although lateral inhibition appears more generally across biology, for instance, to increase sensory perception in neurons [33]. Often, Delta-Notch signaling is combined with some initial prepatterning to create highly ordered structures [10]. However, even from an initial population of identical cells, Delta-Notch signaling will sort cells into two states through lateral inhibition and stochasticity in intracellular dynamics. A two-dimensional (2D) tissue of identical cells can then form a "salt and pepper" pattern of the two cell states under Delta-Notch dynamics.

Notch and Delta are transmembrane proteins located on the surface of cells. When a Delta protein in cell A binds to a Notch receptor on cell B, cleavage of Notch is triggered, releasing the Notch intracellular domain. In turn, the Notch intracellular domain translocates to the nucleus, where it leads to transcriptional changes within cell B. Having received this signal, cell B suppresses the production of Delta [8]. In this way, a cell that is expressing Delta will inhibit its neighbors from similarly expressing Delta, thus forming a laterally inhibiting system [Fig. 1(c)]. The two final states are cells with high Notch and low Delta expression, and cells with high Delta expression. To model the full complexities of a specific

Delta-Notch patterning system, one would have to account for both the Notch and Delta proteins, their mRNA transcripts, as well as a number of transcription factors and their mRNA transcripts, like Scute and E(spl)m3-HLH. Indeed, there are many existing models of Delta-Notch dynamics that capture varying levels of biological details [11,29,34]. However, the key principles of lateral inhibition can be captured with a model where cells have a single internal variable [10], analogous to a reaction coordinate between the two terminal cell states [6].

B. Minimal model of lateral inhibition

Specifically, we study a modified version of the Delta-Notch model in Ref. [10], where instead of a continuous internal variable, our cells have a discrete internal state $u \in \{0, 1, \ldots, N\}$. To model cell-cell communication, each cell has a signal receiver state s where $s = 0$ means no signal is being received and $s = 1$ means that a signal is being received. In reality, the level of signal depends on how many Notch receptors have been recently cleaved and will not be binary. Cells can adjust their internal state stochastically [Fig. 1(b)],

$$u \underset{f^-(u+1,s)}{\overset{f^+(u,s)}{\rightleftharpoons}} u+1, \tag{1}$$

where the transition rates $f^{\pm}(u, s)$ depend on the cell's internal state as well as its signal receiver state. The receiver state can also change stochastically,

$$s = 0 \underset{k^-}{\overset{k^+}{\rightleftharpoons}} s = 1, \tag{2}$$

where k^- is taken to be a fixed parameter, but k^+ is a function of the neighboring cells. Specifically, for cell i, we take

$$k^+ = \sum_j A_{ij}\, g(u_j), \tag{3}$$

where $g(u)$ determines the strength at which a cell sends a signal to its neighbors, A_{ij} is an adjacency matrix where $A_{ij} = 1$ if cells i and j are neighbors at time t, and $A_{ij} = 0$ otherwise. This form could be easily modified so that the value of the adjacency matrix depends on the area of contact between cells, or even an area of contact that changes with time, but for now we assume that all neighbors have equal and fixed contact areas. Finally, we assume that the states $u = 0$ and $u = N$ are absorbing states; once a cell has entered those states it will not leave. Cells can implement an effective threshold and these absorbing states model a level of Notch or Delta which triggers a fate commitment in that cell [8,12].

With the model specified, we want to know if there exists some set of rates $\{f^{\pm}, k^-, g\}$ under which a system of M cells will coordinate to make a target pattern. Eventually all cells will reach an absorbing state, either 0 or N, after which there are no cell state transitions, and we refer to this as a terminal state. We focus on a system of $M = 3$ cells where every cell is in contact with every other cell, and the target pattern is one cell in the N state and two cells in the 0 state (Fig. 1). This is the smallest network that exhibits decentralized symmetry breaking with each cell coordinating with multiple neighbors.

Crucially, every cell has the exact same set of transition rates; if one cell had rates with $f^+ > 0$, $f^- = 0$ and the other two cells had different rates with $f^- > 0$, $f^+ = 0$ then a correct patterning would always be achieved. Instead, we give every cell the same rates and identical initial conditions. If each cell were to stochastically choose an absorbing state without communication, our target pattern would be achieved with probability at most 4/9 (Appendix A). To reliably reach the target pattern, cells must coordinate with each other through signaling. In the next section, we explore the optimal strategies for reaching the target state, before quantifying the information transferred between cells executing an optimal strategy in Sec. IV.

III. SPEED-ACCURACY TRADE-OFF FOR OPTIMAL PATTERNING

In our system, cells can receive a signal (through s), control their internal state (through $f^{\pm}$), and send a signal to neighboring cells (through g). With these essential ingredients for self-organization, our first question is whether regions of the parameter space, $p = \{f^{\pm}, k^-, g\}$, can achieve patterning with a small error, ϵ, which we define as the probability of our system ending up in a nontarget terminal state. Scaling by the fastest timescale, so that $f^{\pm}, k^-, g \leqslant 1$, our model allows for an arbitrarily small error rate (Appendix C), although this comes at the price of taking infinitely long to reach the terminal states (Appendix B). This trade-off is generic, occurring for any value of $M \geqslant 3$ and any adjacency matrix (Appendix B). Therefore, a more biologically motivated optimization problem is to specify the degree of error that is permitted, ϵ_{tol}, and minimize the average time to reach the terminal states, τ, under the constraint that $\epsilon \leqslant \epsilon_{\text{tol}}$. We note the same region of the speed-error plane is accessible whether one constrains $\epsilon \leqslant \epsilon_{\text{tol}}$ and minimizes τ or if one constraints $\tau \leqslant \tau_{\max}$ and minimizes ϵ.

The level of permitted error, ϵ_{tol}, in a fate specification depends on its context within development, and whether any mistake can be corrected or compensated for later on. For instance, ABp specification via Notch signaling is critical for normal development in *C. elegans* [35], whereas when Notch signaling specifies sensory organ precursor cells in *Drosophila* pupal abdomen, errors can be compensated by later cell rearrangements and in any case a single misspecification is not critical to survival [9]. Ideally, we want to examine how the minimum average time to reach a terminal state changes as a function of the error constraint ϵ_{tol}, a Pareto front that, barring additional considerations such as evolvability [36] or thermodynamic cost of transcription, translation, and signaling [37], evolutionarily optimized systems should be near. Similar trade-off structures appear for speed and dissipation in transcription [37], for fluctuations in biochemical reaction networks [38–40], and for dissipation, speed, and accuracy for synchronized oscillators [41,42].

Mathematically, our model is a continuous-time Markov chain (CTMC) for the full state of the system $x_\alpha = ((u_1, s_1), \ldots, (u_M, s_M))$, tracking both the internal and receiver states of every cell. A probability distribution over the states evolves according to the master equation,

$$\frac{\mathrm{d}}{\mathrm{d}t} P_t(x_\alpha) = \sum_\beta Q_{\alpha\beta} P_t(x_\beta), \tag{4}$$

where $P_t(x_\alpha)$ is the probability of the system being in state x_α at time t, $Q_{\alpha\beta}$ is the rate at which x_β transitions to x_α for $\alpha \neq \beta$, and $Q_{\alpha\alpha} = -\sum_{\beta\neq\alpha} Q_{\beta\alpha}$. We can solve this master equation directly, or simulate realizations of the process with the Gillespie algorithm. Finding the error rate ϵ or the average time to terminal states, τ, can be done directly from the rate matrix Q, by posing them as first-passage-like problems. To do so, we first define τ_α to be the expected time to reach a terminal state starting from state x_α and ϵ_α to be the probability that, starting from x_α, the eventual terminal state does not achieve the target pattern. Calling $\mathcal{T}^G$ the set of terminal states that achieve the target pattern and $\mathcal{T}^B$ the set of terminal states that do not achieve the target pattern, then $\tau_\alpha = 0$ for $\alpha \in \mathcal{T}^G \cup \mathcal{T}^B$, $\epsilon_\alpha = 0$ for $\alpha \in \mathcal{T}^G$, and $\epsilon_\alpha = 1$ for $\alpha \in \mathcal{T}^B$. For the remaining $\alpha \notin \mathcal{T}^G \cup \mathcal{T}^B$, it follows that

$$\tau_\alpha = -\frac{1}{Q_{\alpha\alpha}} - \sum_{\beta\neq\alpha} \frac{Q_{\beta\alpha}}{Q_{\alpha\alpha}} \tau_\beta, \tag{5}$$

$$\epsilon_\alpha = -\sum_{\beta\neq\alpha} \frac{Q_{\beta\alpha}}{Q_{\alpha\alpha}} \epsilon_\beta, \tag{6}$$

and these linear equations can be solved to find τ_α and ϵ_α. Since the rate matrix depends on the parameters $p = \{f^\pm, k^-, g\}$, and τ_α and ϵ_α depend on the rate matrix, we have that $\tau_\alpha = \tau_\alpha(p)$, $\epsilon_\alpha = \epsilon_\alpha(p)$. Supposing that we initialize the system in state α^*, we can now pose the minimization problem as

$$\min_p \tau_{\alpha^*}(p) \quad \text{such that} \quad \epsilon_{\alpha^*}(p) \leqslant \epsilon_{\text{tol}}, \tag{7}$$

where, after minimization, we find both the optimal τ along with a set of parameters p that generate it. As the cells have a discrete internal state, there are only finitely many parameters that we are optimizing over, specifically $4(N-1)$ for the $f^\pm$, 1 for k^-, and $N+1$ for g. The optimization problem is equivalent to a quadratic problem with quadratic constraints (Supplemental Material (SM) Sec. II [43]), a generically hard problem [44]. In our numerical parameter scans, however, we consistently find that solutions converge to a small number of local minima from different initializations, whether solving as interior point optimization or through gradient descent (SM Sec. II [43]).

Minimizing across many values of ϵ_{tol}, with $N = 6$ internal states and $\alpha^* = ((1,0),(1,0),(1,0))$, identifies the optimal trade-off curve between error and average time to reach the terminal state [Fig. 2(d)]. Without penalizing error, by setting $\epsilon_{\text{tol}} = 1$, all cells immediately head toward the closest absorbing state, $u = 0$, and consequently never reach a target pattern. In the other limit of extreme precision, $\epsilon_{\text{tol}} \to 0$, it takes increasingly long to reach a terminal state. For intermediate values, the optimal strategy is able to consistently reach a terminal state in a relatively short time compared to the fastest timescale of the system [Fig. 2(d)]. Above the Pareto front, any parameter combination is accessible [Fig. 2(d)]. While our focus in the main text is on the tractable three-cell case, it is possible to explore the speed-accuracy trade-off in larger systems by using a sampling-based gradient descent scheme (SM Sec. II [43]). To illustrate this scheme, we extended the model to seven cells, modeling an asymmetric two-dimensional epithelial region with nontrivial neighborhood structure, and obtained a Pareto front comparable to the three-cell case (SM Fig. S2 [43]).

Generally, the optimal strategy appears to be one of cells stochastically increasing their internal state; the first cell to leave the initial state signals to its neighbors, who decrease their internal state [Figs. 2(a)–2(c)]. These strategies often implement a sharp change in $f^\pm$ as a function of internal state and receiver state and g as a function of internal state [Fig. 2(c)]. Such sharp responses are a common feature of gene regulatory networks and so this feature does not represent an unphysical aspect of the model. The nonmonotonicity of the rates is also plausible given that we are not modeling a single gene but the output of a gene regulatory network which can be highly nonlinear. In any case, constraining the rates to be monotonic has minimal impact on the Pareto front. These strategies can be mathematically interpreted as each cell, for some given receiver state, navigating an effective energy landscape [45], analogous to a Waddington landscape [16] (SM Sec. II [43]), as shown in Fig. 2(e). From the landscape interpretation, we see that for small error rate, such as $\epsilon = 0.02$, cells that are not receiving a signal are prevented from reaching $u = 0$ by an effective energy barrier, and will slowly increase their internal state. Once a cell receives a signal, it rapidly descends to the inhibited state [Fig. 2(e)]. For larger error rates, the landscape flattens, decreasing the energy barrier for fate commitment [Fig 2(e)]. In this case, cells with a large internal state that are receiving a signal can still advance toward the inhibitor state. We observe qualitatively similar optimal strategies for systems with a larger number of cells (SM Fig. S2 [43]).

This strategy, of a single cell stochastically increasing its internal state before signaling to its neighbors [Fig. 2(a)], differs qualitatively from the model in Ref. [10], where all cells increase their internal state concomitantly before one cell eventually becomes dominant and inhibits its neighbors. Both concomitant [10] and nonconcomitant [9] dynamics have been observed in recent live imaging of Delta-Notch patterning. In particular, a recent experiment measured Scute expression, a transcription factor that upregulates Delta, while sensory organ precursor (SOP) specification was occurring in the *Drosophila* dorsal histoblast and saw minimal concomitant increase in Scute [9]. Scute expression appeared to increase in future SOP cells while remaining low in non-SOP cells, which then began to increasingly express Notch [9].

IV. DYNAMIC INFORMATION TRANSFER

Successful collective self-organization requires cells to exchange information. In our system, in the absence of communication, cells can only reach the target pattern with a probability of at most 4/9 (Appendix A). With communication, however, they can find the correct pattern to arbitrary accuracy. In this section, we quantify the information transferred between cells using techniques of information theory [1,32,46,47]. The central object is the mutual information,

$$I(X;Y) = \mathbb{E}\left[\log \frac{P(X,Y)}{P(X)P(Y)}\right], \tag{8}$$

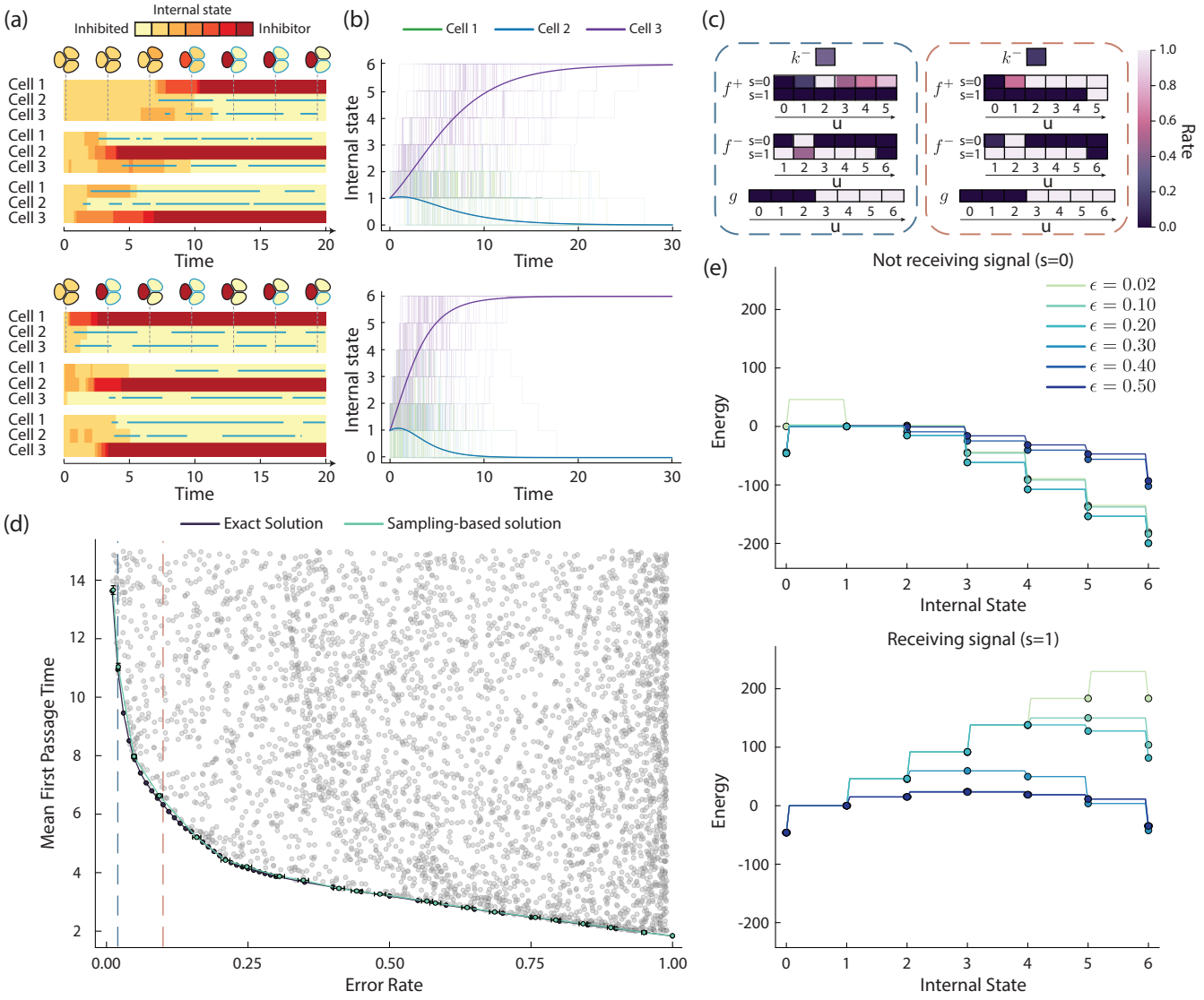

FIG. 2. Self-organizing model is constrained by a speed-accuracy trade-off. (a) Kymographs showing example trajectories from optimized system with error rate $\epsilon = 0.02$ (top) and $\epsilon = 0.1$ (bottom), illustrating how one cell advances stochastically toward the inhibitor state and signals to its neighbors. Color represents internal state; blue line represents a cell receiving a signal. The corresponding multicellular state for select time points is illustrated above for the first kymograph. Throughout this figure, a model with $M = 3$ cells with $N = 6$ internal states is shown. (b) Average internal state for each cell conditioned on a successful trajectory where the inhibitor state was reached by cell 3, for $\epsilon = 0.02$ (top) and $\epsilon = 0.1$ (bottom). Solid lines show the exact average; faint lines show 100 stochastic realizations. (c) Optimized model parameters for $\epsilon = 0.02$ (left) and $\epsilon = 0.1$ (right). (d) Optimizing the mean first-passage time over the set of model parameters p, while constraining the error, finds a speed-accuracy trade-off curve that all models are bounded by (black points). The same trade-off curve obtained using our sampling-based approach (teal points) was computed using 10 000 Gillespie simulations, with error bars showing 1.96× standard error. Additionally, 5000 randomly sampled parameter values (gray points) are shown, demonstrating that the entire region above the trade-off curve is accessible. Dotted blue and red lines show the error rates $\epsilon = 0.02$ and $\epsilon = 0.1$, respectively. (e) Effective "energy landscape" for internal state dynamics are shown for a cell that is not receiving a signal ($s = 0$, top) and a cell that is receiving a signal ($s = 1$, bottom). Points represent the "energy" level of each state and lines represent the height of the effective energy barrier between states (SM Sec. II [43]).

where X and Y are random variables, $P(X)$ and $P(Y)$ are their marginal distributions, and the expectation is taken over $P(X, Y)$, their joint distribution. The mutual information quantifies how much information one gains about X upon seeing Y or vice versa, and is a natural measure of the information shared between X and Y [47,48]. We could use the terminal state of the system to quantify the information shared between neighboring cells in the final pattern, as has been explored recently [1]. However, as we will see, the final state of the system does not quantify all the dynamic information shared between cells during patterning. Quantifying dynamic information requires computing the mutual information between trajectories, $I(X_0^T; Y_0^T)$, where X_0^T denotes the trajectory of a time-dependent variable $X(t)$ for $0 \leqslant t \leqslant T$. If $X(t) = u_1(t)$ and $Y(t) = u_2(t)$, then $I(X_0^T; Y_0^T)$ quantifies the total information that is shared between cell 1 and cell 2 up until time T.

We also make use of the transfer entropy rate, defined as

$$\dot{\mathcal{T}}^{Y\to X}(t) = \lim_{dt\to 0}\frac{1}{dt}\mathbb{E}\left[\log\frac{P_{t+dt}\left[X(t+dt)|X_0^t, Y_0^t\right]}{P_{t+dt}\left[X(t+dt)|X_0^t\right]}\right], \quad (9)$$

which quantifies the directed information transfer rate from Y to X and, provided X and Y cannot simultaneously change [32], is related to the mutual information by

$$I\left(X_0^T; Y_0^T\right) = \int_0^T [\dot{\mathcal{T}}^{Y\to X}(t) + \dot{\mathcal{T}}^{X\to Y}(t)]\, dt. \quad (10)$$

In this section, we measure the magnitude of the information flow, and, by computing conditional transfer entropy rates, we explore how information is transferred from inhibitor to inhibited cells and vice versa, as well as between inhibited cells.

A. Feedback in lateral inhibition makes information calculations challenging

It is well established theoretically that gene regulatory networks can process information, and that the mutual information between the network's inputs and its outputs can be quantified [37,48–52]. Information has been quantified in experimental systems, both directly from trajectories [53–55] and from building detailed models that are fit to experimental data [46,56]. A handful of developmental problems have been studied from the perspective of information theory; for instance, the information contained in the *Drosophila* gap gene pattern was experimentally quantified [4,47,57], and it has been proposed that the gene regulatory networks optimally create an information-rich pattern [58]. In this existing literature, the signal is taken as exogenous; a cell's response to a signal does not impact the signal's future values. While this assumption is often appropriate, such as in ascidian neural induction [18,19], it need not hold in general, and indeed does not hold for lateral inhibition. In our model, cells are both sending and receiving signals and the signals they receive affect the future signals they will send. Any information theory analysis of our model system must take this into account.

In addition to this feedback, other common approximations that simplify the computation of information-theoretic quantities do not hold for a laterally inhibiting system. As we have seen, such systems are time dependent and cannot be approximated as a series of static input and output relationships. Moreover, the multimodal nature of cell fate transitions means that we cannot approximate the dynamics as a unimodal Gaussian process, for which computations are more tractable [59,60]. It is, in theory, possible to simulate our system with the Gillespie algorithm and, treating the simulated data like experimental data, attempt to directly estimate the mutual information. However, direct estimation of mutual information remains challenging due to the high dimensionality of the space of trajectories and such estimators do not take advantage of any of the known structure of our model.

B. Reframing the model as a stochastic reaction network enables tractable information calculations

Computing the mutual information between trajectories is difficult even when the underlying stochastic model is known. To see this difficulty, suppose we wanted to compute a mutual information, $I(X_0^T; Y_0^T)$, where $X(t)$ could be $X(t) = u_1(t)$ or $X(t) = [u_1(t), u_2(t)]$, and similarly for $Y(t)$. While the path measure of a CTMC has a closed-form expression, working out the mutual information requires computing marginal path measures, such as $P(X_0^T)$, which are intractable analytically. Recently, Monte Carlo sampling techniques that take advantage of the known model structure have been used to estimate otherwise intractable terms in mutual-information-like computations [61,62]. In these approaches an outer expectation is estimated by Monte Carlo sampling but once a sample has been drawn, actually computing the corresponding value of the integrand requires another round of Monte Carlo sampling. While advanced sampling techniques like these approaches may be required for many problems, here we can exactly compute the integrand in Eq. (10) following Ref. [32]. There, the authors consider a stochastic reaction network with K reaction channels, n chemical species $Z_1, \ldots, Z_n$, with vector $Z(t)$ recording the copy number of each species at time t, and the kth reaction occurring at a rate $\lambda_k(Z(t))$. If $X(t)$ and $Y(t)$ are variables (or disjoint subsets of variables) in $Z(t)$ that do not change simultaneously, then

$$I\left(X_0^T; Y_0^T\right) = \mathbb{E}\left(\sum_{k\in R_X}\int_0^T \log\frac{\lambda_k^{XY}(s)}{\lambda_k^X(s)} dN_k(s) + \sum_{k\in R_Y}\int_0^T \log\frac{\lambda_k^{XY}(s)}{\lambda_k^Y(s)} dN_k(s)\right), \quad (11)$$

where R_A denotes the set of reactions that involve a change in A, $dN_k(t)$ is the increment of $N_k(t)$ which counts the number of times reaction k has occurred, and $\lambda_k^A = \mathbb{E}[\lambda_k(Z(t))|A_0^T]$ is the expected rate of the kth reaction given A_0^T. The first and second terms in the expectation correspond to the transfer entropies $\mathcal{T}^{Y\to X}(T)$ and $\mathcal{T}^{X\to Y}(T)$, respectively [32], where throughout we define the transfer entropy to be $\mathcal{T}^{A\to B}(T) = \int_0^T \dot{\mathcal{T}}^{A\to B}(t)dt$. Note that Eq. (11) appears as Eq. A6 in Ref. [32], although here we have neglected the integrals with expectation zero.

To actually compute the integrand, we need to compute λ_k^A, which requires knowing the conditional distribution $P(\bar{Z}(t) = \bar{z}|A_0^T)$, where $\bar{Z}(t)$ is a vector that tracks all molecular abundances except those in A. This conditional distribution obeys a stochastic differential equation known as the *filtering equation* [32,63] (SM Sec. III [43]). Keeping track of the probability for every possible count of the latent species $\bar{z}$ is typically not practical in chemical reaction networks, and the filtering equation is instead used to construct moment closure approximations [32,63].

We can interpret our model, Eqs. (1)–(3), as a stochastic reaction network, albeit an unusual one with nonlinear reaction rates and with the copy number of each species bounded above by a finite number. This finiteness of copy number means that we need not approximate the filtering equation, and instead can solve it exactly. To compute the mutual information, we use Monte Carlo sampling to approximate the outer expectation in Eq. (11), but for each sample we exactly compute the term inside the expectation by directly solving the filtering equation (SM Sec. III [43]).

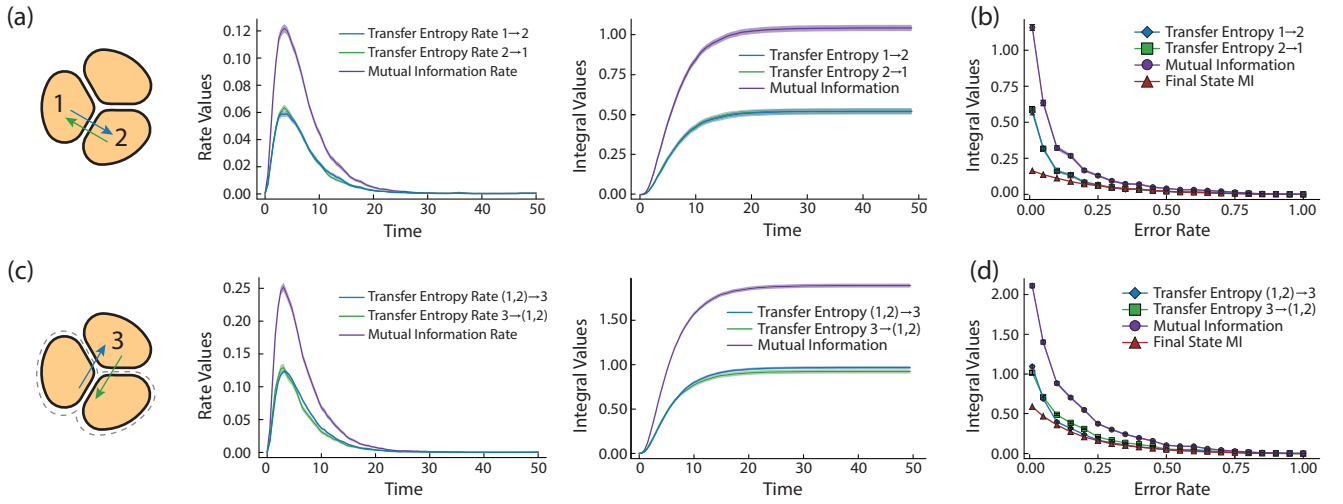

FIG. 3. Quantifying information shared between cell trajectories finds a nonmonotonic rate of information transfer as well as a greater amount of information shared between trajectories than final states alone. (a) For any pair of cells, we can compute the transfer entropy (TE) rates in both directions, $\dot{\mathcal{T}}^{u_1\to u_2}$, $\dot{\mathcal{T}}^{u_2\to u_1}$, the mutual information (MI) rate $\dot{I} = \dot{\mathcal{T}}^{u_1\to u_2} + \dot{\mathcal{T}}^{u_2\to u_1}$ (left), as well as the corresponding integrals (right), shown here for the optimized model with $\epsilon = 0.02$. (b) Total information transferred between any pair of cells for optimized models as the error rate is varied. Each value represents the limiting mutual information or transfer entropy for a given model at large time. Instantaneous mutual information shared between the final states of both cells is also shown. (c) Transfer entropy rates between cell 3 and the remaining cells, $\dot{\mathcal{T}}^{u_3\to[u_1,u_2]}$, the converse, $\dot{\mathcal{T}}^{[u_1,u_2]\to u_3}$, as well as the mutual information rate (left) along with the corresponding integrals (right), shown here for the optimized model with $\epsilon = 0.02$. (d) Total transfer entropy from one cell to the remaining cells and the converse, as well as the mutual information for optimized models as the allowed error rate is varied. Instantaneous mutual information shared between the final state of cell 3 and the remaining cells is also shown. Throughout, each computation is averaged over $n = 10\,000$ Monte Carlo samples and shaded regions show $\pm 1.96\times$ standard error. Integral values are expressed in units of nats, while rates are given in nats per unit time.

Similarly to the information $I(X_0^T; Y_0^T)$, we can compute an information rate $\frac{dI(X_0^t;Y_0^t)}{dt}$, as well as transfer entropy rates, $\dot{\mathcal{T}}^{X\to Y}(t)$ and $\dot{\mathcal{T}}^{Y\to X}(t)$, representing the directed rate of information transfer from X to Y and vice versa. See the SM [43] for additional numerical details.

C. Quantifying information transfer rates numerically

We are now in a position to quantify information flows in our system. For the optimized model with error $\epsilon = 0.02$, and taking $X = u_1$, $Y = u_2$, we find that there is a sharp increase in the rate of mutual information which then sharply decreases [Fig. 3(a)]. Each cell will eventually reach an absorbing state after which no further "reactions" involving X or Y occur, and hence that trajectory makes no further contributions to the mutual information in Eq. (11). The probability that a cell has not reached an absorbing state decays exponentially with time and, indeed, the total mutual information between the trajectories of u_1 and u_2 clearly asymptotes as $T \to \infty$ [Fig. 3(a)]. Due to the symmetry of the problem, any pair of internal variables will have the same mutual information between their trajectories as any other pair and the transfer entropy between them in both directions will be exactly half the value of the mutual information.

For any given value of the error, we can find the optimal model and compute the exact same information-theoretic quantities as we have above for $\epsilon = 0.02$. We find that the total information exchanged between two trajectories decreases monotonically as the allowed error increases, reaching around zero at roughly the point when a zero-communication strategy is possible [Fig. 3(b)]. In the opposite limit of $\epsilon \to 0$, it is possible for the trajectory of cells to share an arbitrarily large amount of information with each other (Appendix D).

In addition to looking at pairs of cells, we can take one cell and look at the information it has about the remaining cells, or $X = u_1$, $Y = [u_2, u_3]$. We similarly see a sharp increase in the information shared between them before this curve asymptotes, at around 1.89 nats for $\epsilon = 0.02$ [Fig. 3(c)]. Interestingly, this quantity greatly exceeds the 0.693 nats (1 bit) that can be shared between a cell and a fixed binary external signal. Additionally, the transfer entropy also exceeds 1 bit, and is asymmetric with each cell receiving more information from its neighbors than it sends [Fig. 3(c)]. This finding shows that, whether through signaling timing or repeated signal activation, the binarized receiver state is capable of transferring more than 1 bit of information. Similarly to the pair of cells, the mutual information between one cell and the remaining cells decreases monotonically as the allowed error increases [Fig. 3(d)].

D. Successful patterning displays directed communication between all cells

Due to the inherent symmetry of our system, each cell is initially equally likely to be the inhibitor cell, and this symmetry obscures the way information is transferred. For instance, suppose that information only flowed from the inhibitor cell to the inhibited cells. The naive transfer entropy calculation would still find every pairwise transfer entropy rate to be equal, say, $\dot{\mathcal{T}}^{1\to 2}(t) = \dot{\mathcal{T}}^{2\to 1}(t)$, since this quantity

is computed over trajectories where cell 1 is the inhibitor and trajectories where cell 2 is the inhibitor. Instead, we would like to decompose our information flows by somehow removing the symmetry that any one of the cells is equally likely to be the inhibitor. In systems with feedback, such as this one, such decompositions can violate the data processing inequality or subtly introduce fictitious dependencies, and thus require careful interpretation. With those caveats in mind, we can define transfer entropy rates conditioned on some terminal event, for instance, $\mathcal{T}^{1\to 2|Z}(t)$, where Z is the event that terminal state is $u_1 = N$, $u_2 = u_3 = 0$ is reached. In general, for the event Z that some final absorbing state is ultimately reached, conditioning the CTMC on Z results in another CTMC. Specifically, using Doob's h transform [64], the rate matrix of the conditioned system is

$$Q_{\alpha\beta|Z} = Q_{\alpha\beta}\frac{P(Z|\alpha)}{P(Z|\beta)}, \tag{12}$$

for $\alpha \neq \beta$ and $Q_{\beta\beta|Z} = -\sum_{\alpha\neq\beta} Q_{\alpha\beta|Z}$, where $P(Z|\alpha)$ is the probability of the event Z given the system is in state α. Quantities of the form $P(Z|\alpha)$ can be computed in the same way we compute ϵ_α [Eq. (5)], after which we can use this new conditioned CTMC to compute information-theoretic quantities exactly as before. We find that the conditioned transfer entropy rates are significantly higher from the inhibitor cell to the inhibited cells than in the reverse direction [Figs. 4(a) and 4(c)]. Nevertheless, the flow of information goes both ways with a nonzero conditioned transfer rate from the inhibited states to the inhibitor state [Fig. 4(c)]. Furthermore, there is a nonzero conditioned transfer rate between the two inhibited cells (Appendix E). Exploring these information-theoretic quantities highlights the complex picture of information flows in a self-organizing system.

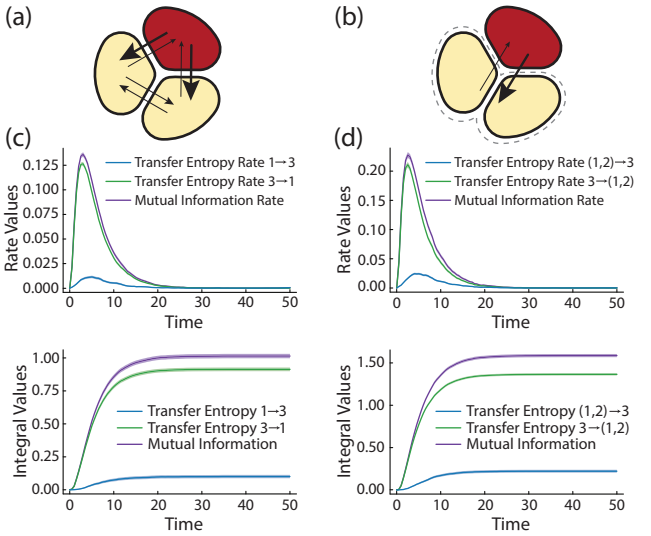

FIG. 4. Conditioning on final cell fates reveals persistent and asymmetric information transfer between cells. Here, all information-theoretic quantities are computed for a process conditioned on the terminal state of cell 3 ending up as an inhibitor $u_3 = N$, while cells 1 and 2 are inhibited, $u_1 = u_2 = 0$. [(a), (b)] Schematic where arrow thickness is proportional to the long-time conditional transfer entropy between (a) pairs of cells and (b) the inhibitor cell and the remaining pair. (c) Conditioned mutual information and transfer entropy rates between cell 2 and cell 3 (top) as well as their corresponding integrals (bottom). Interestingly, the conditional information transferred from the inhibited cells is small but positive. (d) Conditional mutual information and transfer entropy rates between cell 3 and and its neighbors (top), alongside their corresponding integrals (bottom). Throughout, each computation is averaged over $n = 10\,000$ Monte Carlo samples, and the shaded regions show $1.96\times$ standard error. Integral values are expressed in units of nats, while rates are given in nats per unit time.

V. INSTANTANEOUS VERSUS DYNAMIC INFORMATION TRANSFER

To completely quantify the information shared between cells and to account for situations where a change in one cell only affects another cell at a later time, information quantities should be computed between entire trajectories, as we have done so far. However, while it is possible to compute these quantities exactly in our model, computing the mutual information between trajectories directly from biological data remains impractical due to the high-dimensional space which trajectories inhabit. A more practical approach is to compute information-theoretic quantities using only the instantaneous state of the system, rather than full trajectories. In this section, we explore what can be learned from instantaneous information quantities alone.

Consider the mutual information between the internal state of two cells at some fixed time t, $I(u_1(t);u_2(t))$, which we refer to as the instantaneous mutual information. Due to the data processing inequality, the instantaneous mutual information is always less than the mutual information between the trajectories up until that point, or $I(X_0^t;Y_0^t) \geqslant I(X(t);Y(t))$. A related quantity called correlational information was recently proposed [1], defined as

$$\mathrm{C}(t) = \frac{1}{M}\mathbb{E}\left[\log\frac{P(u_1(t),\ldots,u_M(t))}{\prod_{i=1}^{M}P_i(u_i(t))}\right], \tag{13}$$

where, in our system, all cells are equivalent and hence the marginal probability P_i is the same for every cell. Note that $M \times \mathrm{C}$ is sometimes referred to as the multi-information or the total correlation of the random variable $(u_1(t),\ldots,u_M(t))$ [65]. Analogously to how we compute mutual information between trajectories, we can compute a trajectory version of correlational information (SM Sec. III [43]).

Plotting the instantaneous mutual information for a particular choice of $\epsilon = 0.02$, we find that the instantaneous information is indeed less than the full dynamic information by a factor of around 4–6 (Fig. 5). In the long-time limit, $t \to \infty$, the instantaneous mutual information between a pair of cells is around 0.156 nats, or around 15.3% of the total information shared between the full trajectories. The final-state instantaneous mutual information is smaller than the dynamic mutual information for all values of ϵ, with the largest difference occurring at small ϵ [Figs. 3(b) and 3(c)]. In fact, since the final state can only take one of two values, $u_1(t\to\infty) \in \{0, N\}$, we have $I(u_1(t\to\infty);u_2(t\to\infty)) \leqslant$

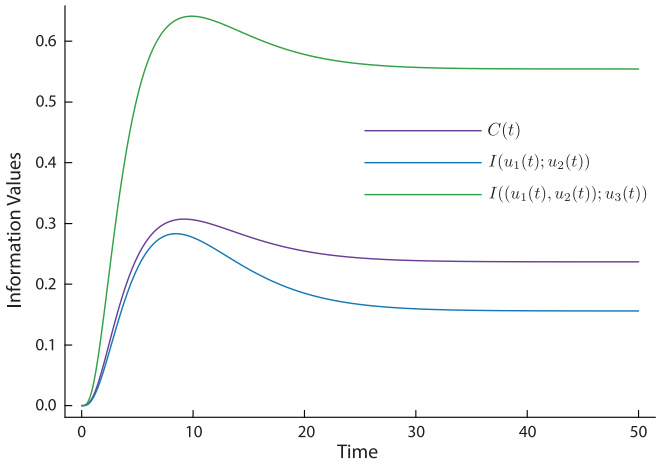

FIG. 5. Instantaneous information quantities can be nonmonotonic in time during self-organization. The information quantities, C, $I(u_1(t);u_2(t))$, and $I([u_1(t),u_2(t)];u_3(t))$, are computed from the instantaneous state of the system at time t, shown here for $\epsilon = 0.02$. All quantities are expressed in units of nats.

$H(u_1(t \to \infty)) \leqslant \log 2$. In contrast, the dynamic information shared between cells can be unbounded in the $\epsilon \to 0$ limit (Appendix D).

Intriguingly, the shape of the instantaneous mutual information curve, as well as the correlational information curve, is nonmonotonic, exhibiting a local maxima at a finite time (Fig. 5). Heuristically, we can understand this through the following argument. Suppose after some long time you learn that one cell is at state N. You can deduce that its neighbors are likely to be in state 0. If you learn that one cell is at state 0 you deduce that one of its neighbors is likely in state N and the other one is likely in state 0. Now suppose instead that we view the system at some intermediate, but later, time. Upon learning that one cell is in state N or 0, we may make similar deductions as before. However, if the cell is in an intermediate state, we can deduce that the neighboring cells are also likely to be in an intermediate state. Therefore, at the intermediate time, there are more possible states for the neighbors to be in, and hence one can learn more information about the neighbors by knowing the state of a cell than at the final time. This highlights how a careful interpretation of instantaneous mutual information is required, and that a decrease in instantaneous mutual information does not necessarily mean that the system is becoming less coupled or more disordered.

VI. COLLECTIVE AND INDIVIDUAL OPTIMALITY

Through evolutionary selection, development is tuned to robustly generate viable, functional offspring. Many cellular processes, including Delta-Notch signaling, are highly conserved and have been tuned across millions of years of evolution. It is therefore natural to ask questions about optimality, such as how gene regulatory networks can optimally process intracellular information [48–50], how cells can optimally infer an external signal [12,13,66], and how cells can act optimally to control their environment [15]. Similarly, development has been modeled by cells acting as optimal Bayesian agents, seeking to minimize uncertainty about their cellular identity [30]. However, fitness is defined at the level of the organism or even the level of the community, making it challenging to ascribe optimality to any particular component of the system. We find that by optimizing a collective objective, the actions of individual cells need not appear individually optimal. Additionally, we find that the collectively optimal system does not maximize information flow. In this section, we briefly examine how collective optimization results in individual cells appearing to act suboptimally.

A. Local sacrifices enable global gains

Previously, in Eq. (7) we minimized over one set of parameters p that every cell shared. Equivalently, we could have given each cell its own set of parameters p_1, p_2, and p_3, and optimized over all of these together with the constraint that $p_1 = p_2 = p_3$. Fix the parameters of two cells, say, cells 1 and 2, to follow the collective optimum, $p_1^* = p_2^*$, and consider the remaining cell 3. From the perspective of this cell alone, the strategy given by the collectively optimized parameters p_3^* is not optimal. Cell 3 can optimize its parameters to find a new strategy $\tilde{p}_3$ which achieves a faster time to the terminal state with the same error (Fig. 6). However, all cells adopting p_3^* is superior to all cells adopting $\tilde{p}_3$ as in the latter case (i) the resulting system is more error prone, and (ii) at the new error rate, the strategy is far from optimal (Fig. 6). This observation reminds us that for self-organizational problems the objective is a collective one. The actions of a single cell, when treating the remaining cells and signaling environment as an exogenous mean field, may appear suboptimal. Only in the context of the collective problem are cells' actions optimal. Similar trade-offs where individual strategies appear suboptimal yet enable collective coordination have been recently observed at larger biological scales [67].

B. Total information flow is not the objective

A signaling pathway, taken in an isolated cell with an exogenous signal, is often considered as optimally translating the information from the external signal into an internal state [37,48]. However, for a collective, self-interacting system with feedback, where the signal is sent and received by cells, it becomes less clear what optimal signal processing should look like. For a given rate of error, we find suboptimal solutions that both transfer more information between cells than the optimal solution as well as less information (Fig. 7). When the objective is simply to reach the target state as quickly as possible with some allowed error rate, cells do not optimize the information transferred between them.

VII. SELF-ORGANIZED PATTERNING IN *DROSOPHILA* SENSORY ORGAN FORMATION

To explore our theoretical predictions, we examine recent experimental data of Scute expression during sensory organ patterning in the *Drosophila* pupal abdomen, an example of lateral inhibition through Delta-Notch signaling [9]. Directly computing the full mutual information between trajectories remains impractical from finite experimental data due to the high dimensionality of trajectories [62]. Instead, we compute the instantaneous mutual information directly from data

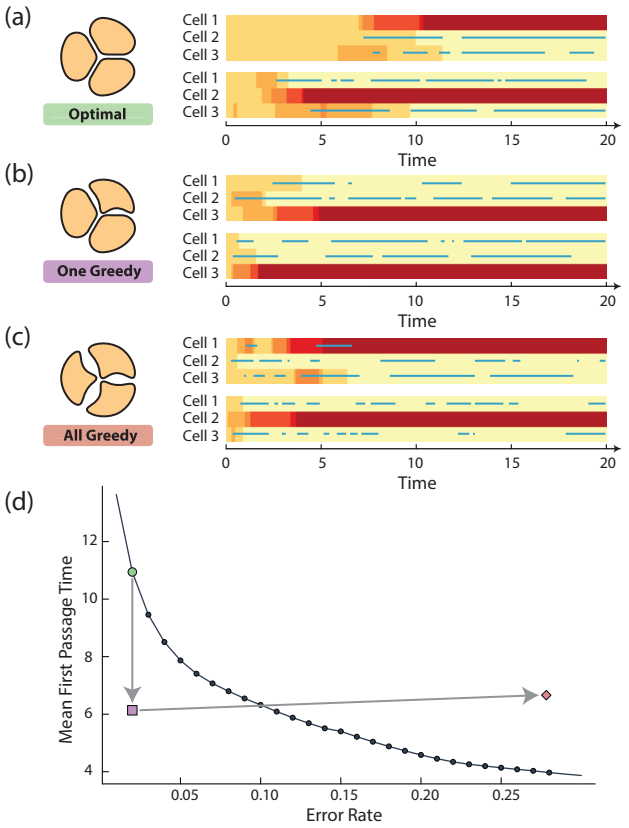

FIG. 6. Collective optimal strategy does not appear locally optimal. Example kymographs for (a) all cells collectively optimized [$\epsilon = 0.02$, as in Fig. 2(a)], (b) optimizing a single cell with the remaining cells fixed, and (c) three copies of the individually optimized cell. (d) Speed-error trade-off plot [from Fig. 2(a)], showing collective optimal strategy (green circle), a single optimized cell with remaining cells fixed (purple square), and three individually optimized cells combined (red diamond).

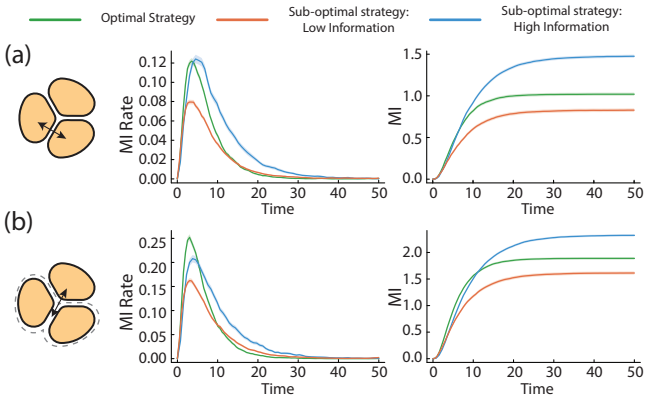

FIG. 7. Collective strategy does not optimize for information flow. For a fixed error rate of $\epsilon = 0.02$, suboptimal strategies that transfer more or less information between cells than the optimal strategy can be found. [(a), (b)] Mutual information rate (middle) and total mutual information (right) transferred between (a) a pair of cells, and (b) a cell and its neighbors, in the collectively optimal system (green), a system that transfers more information for the same error rate (blue), and one that transfers less information (orange). Throughout, each computation is averaged over $n = 10\,000$ Monte Carlo samples, and the shaded regions show $1.96\times$ standard error. Integral values are expressed in units of nats, while rates are given in nats per unit time.

without assuming any particular underlying model. We find that this measure varies nonmonotonically over the course of cellular patterning, consistent with our theoretical analysis in Sec. V.

We analyze live imaging data of sensory organ formation in Drosophila pupal abdomen from Ref. [9]. In these experiments, the transcription factor Scute, along with a nuclear marker, were endogenously tagged with fluorescent reporters. Scute is part of the Delta-Notch feedback loop, it upregulates expression of Delta, and it is indirectly suppressed by the activation of Notch receptors. Thus, Scute serves as a proxy for the "reaction coordinate" between the inhibited state and the inhibitor state, with high Scute expression indicating the inhibitor state and low Scute expression indicating the inhibited state. Moreover, the fluorescence intensity of Scute within a nucleus provides a quantitative readout of its expression for each cell simultaneously. During the imaging period, cells coordinate through Delta-Notch signaling to create a somewhat regular pattern of Delta expressing cells, known asSOP cells [Fig. 8(a)]. Errors where two neighboring cells are both SOPs occur around 10% of the time, but can be corrected through cell rearrangements [9]. In total, three live-imaging experiments of this type were performed. Each experiment was imaged for around 12 h and had around 300 time points each of resolution $660 \times 900 \times 24$ pixels^3 corresponding to a physical region of around $257.4 \times 351 \times 31.9$ μm^3.

A direct application of the mutual information formula is complicated by the fact that cells divide, die, and move within the tissue. Following Ref. [9], we instead take the intensity of Scute in a cell and compare it to the average intensity of its neighbors [Figs. 8(b) and 8(c)]. The neighborhood is determined by taking the three-dimensional Delaunay tessellation of the nuclei centroids and retaining only edges that are shorter than 12 μm (SM Sec. VII [43]). Note that this neighborhood can change in time. We then apply a difference-of-Gaussian filter to the intensities to remove regional variations in background fluorescence, which could erroneously correlate low-Scute-intensity cells and their neighbors (SM Sec. VII [43]). Having processed the data, for each time point we have a series of pairs $(u_i(t), v_i(t))$ where $u_i(t)$ is the intensity of the ith cell, and $v_i(t)$ the mean intensity of its neighbors. To compute the mutual information between u and v, we apply the Kraskov-Stögbauer-Grassberger estimator [68] (SM Sec. IV [43]). Estimating mutual information from finite data is challenging and data from uncorrelated variables can result in a nonzero point estimate. To confirm that our estimate of mutual information is statistically significant, we create a null dataset by randomly reassigning the identity of each neighbor, essentially creating a set of pairs $(u_i(t), v_{\sigma(i)}(t))$ where σ is a random permutation. In this null dataset the marginals are preserved but now u and v are approximately independent. For much of the time when patterning is occurring, the mutual information estimation exceeds 95% of estimations computed from the null data set, demonstrating that there is statistically significant information

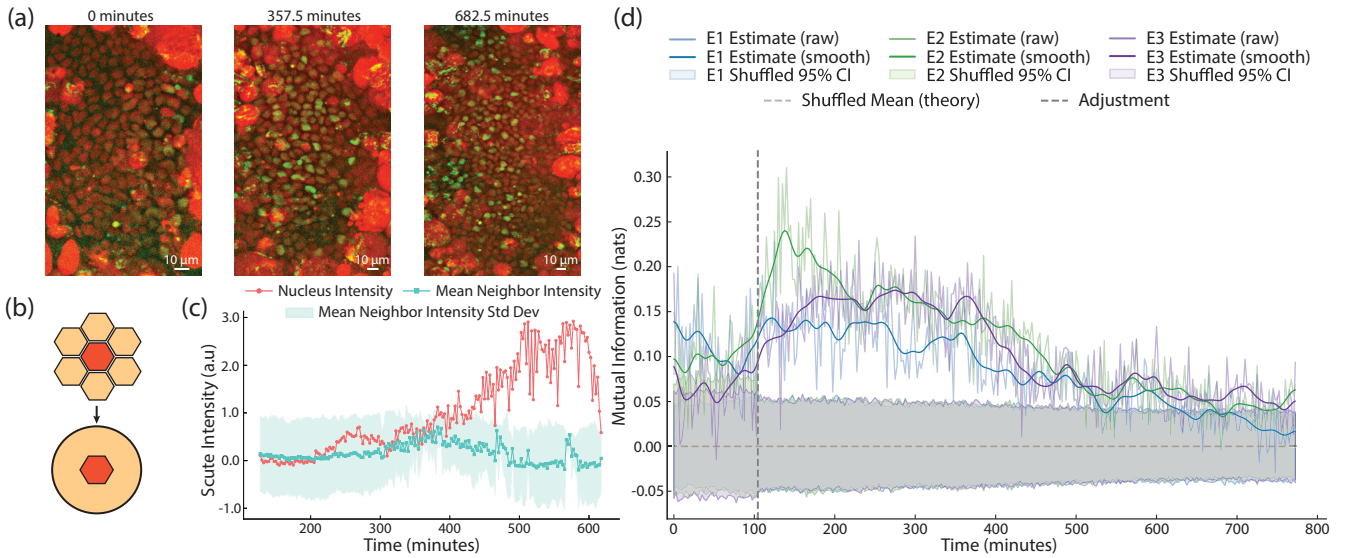

FIG. 8. Instantaneous mutual information between cells for experimental measurements of *Drosophila* pupal abdomen patterning show nonmonotonicity in time. (a) Snapshots of live-imaging experiments from Ref. [9], showing a nuclear marker (red) and Scute intensity (green). Maximum intensity projections of a three-dimensional raw image are shown. Throughout, time is measured from the start of the experiment, around 12 h after puparium formation. (b) We compare the intensity of a cell to the average intensity of its spatial neighbors, which may change with time as cells rearrange and divide. (c) Example Scute intensity trajectory of a future SOP cell compared to its neighbors over the time window when patterning occurs (SM Sec. VII [43]). Scute ultimately decreases once a cell commits to SOP fate. (d) Instantaneous mutual information between cells and their neighbors across time and for three separate experiments (E1–E3). The raw mutual information time series is overlaid with a temporally smoothed trend (SM Sec. VII [43]). Dotted line indicates discontinuity in the imaging region, and number of segmented cells, due to scope adjustment (SM Sec. VII [43]). Shuffling cell-neighbor intensity pairs gives a null dataset with zero true mutual information; numerical confidence intervals (95% shown) show values of estimator applied to null data.

shared between a cell and its neighbors [Fig. 8(d)]. We additionally see that the instantaneous information decreases as patterning occurs and this is observed across all experiments [Fig. 8(d)], a potentially misleading feature of instantaneous information that we also observed in the model in Sec. V.

VIII. DISCUSSION

To explore the underlying principles of decentralized self-organization, we introduced a tractable model of a laterally inhibiting system and demonstrated that it is capable of reaching a target pattern starting from initially identical cells. Our analysis shows that a trade-off exists between patterning accuracy and time taken to pattern, resulting in a Pareto front of optimized solutions with varying error rates. By reframing the model as a stochastic reaction network, we are able to directly compute quantities such as the mutual information and transfer entropies between cell trajectories, revealing how information is transferred between cells. Having optimized for collective patterning speed and error rate, the solutions do not appear to optimize for the flow of information, nor do cells appear individually optimal when considered in isolation. Computing the mutual information between trajectories from data alone remains challenging, and so we explore more tractable information measures, such as the instantaneous mutual information, finding that these quantities may display counterintuitive behavior such as nonmonotonicity in time. Finally, we compute such quantities in experimental measurements of Delta-Notch patterning and once again find that instantaneous quantities can be nonmonotonic in time as patterning occurs, even as the total dynamic information shared between cells strictly increases in time. Our main focus is to extend the analysis of optimal behavior and information transfer beyond the single-cell context to a system of decentralized interacting cells engaged in a self-organization problem.

In development, the starting point of patterning is often inhomogeneous. For example, in addition to Notch signaling, there is also a Delta prepattern in SOP formation in *Drosophila* dorsal thorax [10], allowing for a more regular array of SOP cells. Rather than all cells being initially identical, we could consider them to have some initial prepatterning containing some useful, but not perfect, information. This would still mean that cells need to coordinate to achieve a target pattern, but presumably could do so faster than without the prepattern for the same accuracy requirement. How optimal strategies and information flow change as the information contained in the prepattern change remains an open question for future study.

An assumption of our model is that every cell has exactly the same rates as every other cell. In reality, cells are heterogeneous and will differ somewhat in their initial condition as well as their parameters and hence response to a signal [55,69]. This initial heterogeneity need not represent a prepattern or contain useful information; it could be entirely stochastic. Additional sources of heterogeneity come from the nonregular initial packing of cells, so that different cells experience different, and dynamic, neighborhoods. Even so, heterogeneity may well enhance the patterning ability of our system [9], by breaking the initial symmetry between cells. It

would be interesting to explore how varying amounts of heterogeneity change the optimal speed-accuracy trade-off curve.

We have focused on a minimal patterning motif, but in principle our framework can be extended through the sampling-based gradient descent scheme (SM Sec. II [43]). Such an extension is needed to investigate how the character of the optimized solutions depends on the specific choices of our modeling framework. For instance, our framework has discrete internal states and our analysis is carried out at a fixed, moderate value of N, where stochastic fluctuations are significant. While we do not observe sensitivity to the particular value of N in this regime, we have not systematically explored the $N \to \infty$ limit, or included additional states of signal receiving. Another modification of the model would be to dynamically generate the cell-cell contact topology through a vertex model, allowing for cell movement, neighbor exchange, and tissue deformation to feed back onto the signaling network [70–73]. Similarly, one could incorporate cellular growth, division, and death, all of which are observed in SOP formation in *Drosophila* pupal abdomen [9]. When adding such complexity, however, it becomes challenging if not impossible to numerically solve the master equation or directly compute hitting times and error rates. Nonetheless, the sampling-based approach (SM Sec. II [43]) can, in principle, accommodate these additions as long as the likelihood of each simulation can be computed. The extent to which further physical constraints, such as energetic constraints, robustness, or evolvability, shape the character of the solutions remains an interesting open question.

Although our model reproduces behavior that is observed in real experimental systems, such as the scale of information transfer and nonmonotonicity, it is not intended to be a detailed biological model. Relatedly, due to the complexities of gene regulatory networks, rather than modeling every molecular event, conceptual progress has been made on understanding cell fate transitions by using "gene-free" phenomenological models [5,10,16]. Similarly, simple discrete models have proved powerful for exploring the concepts underlying cell fate patterning without explicitly modeling the complex underlying gene regulatory networks [74–76]. Here, we are able to gain insight into optimal patterning and information flow by studying a simplified phenomenological model. Suppose, instead, that we wish to precisely calculate information-theoretic quantities in a specific biological system. As we have seen, some simple information measures, such as the instantaneous mutual information shared between cells, can be computed directly from data but they can be misleading. For instance, the instantaneous mutual information can decrease while the system patterns. While it is conceivable that alternative instantaneous information measures exist without these specific issues, any instantaneous quantity will necessarily be less insightful than the trajectory-based mutual information. Estimating full trajectory quantities directly from experimental data remains impractical with existing estimators due to the high dimensionality of trajectories, although this remains an active area of research [77]. Seemingly, the most promising approach is to build a detailed model, fit this model to data, and then compute the information transfer within this model [37,46]. While such computations have not yet been attempted in a detailed model with multiple cells and feedback, advances in computing transfer entropy rates [61] might render such computations possible, albeit numerically expensive. With the ability to compute information transfer in multicellular systems, it will be possible to ask precise information-theoretic questions, for example: along which direction is information maximally transferred in? How does information transfer depend on the cell contact topology? In this sense, an understanding of how information propagates in a self-organizing system requires a method to precisely compute information-theoretic quantities.

Throughout, for computational simplicity, we bound all rates above by 1, which essentially constrains the fastest timescale in our system. Therefore, strategies which create a high-accuracy pattern at the cost of taking a long time have a timescale separation between the patterning timescale and the timescale of the internal dynamics or signaling. In other words, accurate self-organized patterning requires transcriptional and signaling dynamics to occur on a faster timescale than patterning. For many developmental processes, this may well be a limiting constraint. The timescale of transcriptional and translational dynamics is on the order of tens of minutes [78], which would constrain accurate self-organized patterning to be on the order of hundreds of minutes, which is indeed a typical patterning timescale. While we prove that such a trade-off exists in our model, we expect that this is a generic effect any cell faces in a system with a noisy mechanism of communication: spend longer acquiring information about the environment or act faster at the risk of making an incorrect patterning choice. Indeed, in a continuous model of lateral inhibition, a separation of timescales between the timescale over which cells commit to an SOP fate and the timescale at which a cell can inhibit its neighbors has been identified as necessary for accurate patterning [34]. Beyond lateral inhibition, there are numerous contexts in developmental biology where the duration of a signal, and not just the strength, determines whether cells commit to a particular fate [79,80]. Perhaps the required timescale separation is one reason why prepatterning and exogenous signals are often used in combination with self-organization: self-organization of identical cells can achieve accurate patterning, but such approaches take inordinately long compared to self-organization with prepatterning.

ACKNOWLEDGMENTS

We are grateful to François Schweisguth for sharing experimental data with us and Minh Son Phan for image processing advice. We thank Henry Mattingly, Matthew Smart, and David Denberg for helpful discussions. A.T. acknowledges the support of the Summer@Simons program. The authors also acknowledge the MIT Office of Research Computing and Data for providing high performance computing resources that have contributed to the research results reported within this paper. This research received support through Schmidt Sciences, LLC (J.D.), the MathWorks Professorship Fund (J.D.).

DATA AVAILABILITY

The data that support the findings of this article are openly available [81].

APPENDIX A: NO COMMUNICATION STRATEGY

Without the ability to communicate, each cell can only adjust its internal state to reach either absorbing state, 0 or N. The specific choice of transition rates affects the time it takes to reach an absorbing state, but whatever the choice of rates there will be a probability q of reaching 0 first and a probability $1-q$ of reaching N first. Since each cell has to have the same transition rates, and without communication the cells are independent, the probability that exactly one cell reaches N and that two cells reach 0 is $3q^2(1-q)$. Maximizing over q, we find $q = 2/3$ and the probability is $4/9$.

APPENDIX B: ERROR BOUNDS THE AVERAGE TIME TO PATTERN

Consider a system of $M \geqslant 3$ cells and a set of good terminal states where for every good terminal state the number of cells in state N, or n_d, is some fixed number, $0 < n_d < M$. It need not be the case that any terminal configuration with n_d cells in state N is a good state, only that all good states have n_d cells in state N. Throughout the text, $M = 3$ and a good terminal state has exactly one cell in state N, or $n_d = 1$. For the more general system, suppose there exists some set of parameters p such that the error rate is ϵ. In this case we can show that the average time taken to reach a terminal state will be at least $M^{-2}(3M+3)^{-1}(1-\epsilon)\epsilon^{-(2N+2)^{-M}}$, demonstrating that precision comes at the cost of time. To prove this outcome, we will need the following lemma.

Lemma 1. If there is a nonzero probability of reaching a good terminal state, there is a nonzero probability of reaching a bad terminal state.

Let $\mathcal{X}$ be the set of *self-sufficient* single-cell states: those from which the cell has a nonzero probability of reaching an absorbing state (0 or N) without receiving any further signals. This could include states where the signal-receiving state must turn off to reach an absorbing state. The absorbing states are trivially included in this set. We need not assume anything about the cell-cell adjacency matrix although it suffices to prove the lemma for a connected graph and apply the result to the disconnected components of a general adjacency matrix.

Case 1: $(1,0) \in \mathcal{X}$. Then each cell can, with nonzero probability, reach an absorbing state without ever receiving any signal from its initial state. Hence, there is a path with nonzero probability through which all cells can reach the same absorbing state, which would result in a bad terminal state.

Case 2: $(1,0) \notin \mathcal{X}$. In this case, every cell must receive a signal to progress to an absorbing state. We can trace the steps a cell takes to go from (1,0) to when it first enters $\mathcal{X}$. In doing so, it must be possible for a cell to reach a state $(w,0)$ where $g(w) > 0$ before the cell receives a signal. If not, then no cell could ever receive a signal. Also, there must be a self-sufficient state that is reachable from (1,0), $(v,s) \in \mathcal{X}$, for which $g(v) > 0$, because a successful trajectory starts with no cells being in $\mathcal{X}$, and ends with M cells being in $\mathcal{X}$. Since only one cell state changes at a time, there must be a point at which $M-1$ cells occupy a state in $\mathcal{X}$ and one does not. To be successful, this final cell needs to receive a signal to enter $\mathcal{X}$ which requires a nonzero g from one of the cells already in $\mathcal{X}$.

Now consider the following finite sequence of steps:

(i) Move cell 1 into the state $(w,0)$ with $g(w) > 0$.

(ii) Move the remaining cells, starting with the neighbors of cell 1, into the self-sufficient state (v,s) with $g(v) > 0$.

(iii) Move cell 1 to (N,s) (the exact receiver state does not matter).

(iv) Now move the remaining cells into the absorbing state N if it is accessible from (v,s), else move them to 0.

If state N is accessible from (v,s), or if $n_d \neq 1$, then a path is produced with nonzero probability that reaches a bad terminal state. If N is not accessible from (v,s) and $n_d = 1$, then consider the following modified sequence of steps:

(i) Move cell 1 into the state $(w,0)$ with $g(w) > 0$.

(ii) Move $M-2$ of the remaining $M-1$ cells, starting with the neighbors of cell 1, into the self-sufficient state (v,s) with $g(v) > 0$.

(iii) Move the final cell to (N,s) (the exact receiver state does not matter).

(iv) Move cell 1 to (N,s) (the exact receiver state does not matter).

(v) Now move the remaining cells into the absorbing state 0.

The final state has two cells in state N and hence this represents a path to a bad terminal state with nonzero probability.

Proposition 1. For an M-cell system with error rate $\epsilon < 1$, the average time to reach the terminal states, τ, satisfies $\tau \geqslant M^{-2}(3M+3)^{-1}(1-\epsilon)\epsilon^{-(2N+2)^{-M}}$.

One can interpret a CTMC as a Markov chain for the sequence of states, along with a set of residence times drawn from an exponential distribution. Specifically, we can write the probability of making a particular transition from $j \to i$, given that we are in state j as $\mathbb{P}(j \to i|j) = W_{ij}/\sum_{k\neq j} W_{kj}$. Since, in our system, only $3M$ transitions are possible from any given state, and each transition rate is bounded by 1, we can conclude that $\mathbb{P}(j \to i|j) \geqslant W_{ij}/3M$. For a path $\mathcal{P} = \{\alpha_1 \to \alpha_2 \to \cdots \to \alpha_{l+1}\}$, we have that

$$\begin{aligned}\mathbb{P}(\mathcal{P}|\alpha_1) &\geqslant \prod_{k=1}^{l}\left(\frac{W_{\alpha_{k+1}\alpha_k}}{3M}\right)\\ &\geqslant \left[\min_{1\leqslant k\leqslant l} W_{\alpha_{k+1}\alpha_k}/3M\right]^l,\end{aligned} \tag{B1}$$

where, by considering solely the Markov chain path, we have effectively marginalized over the possible waiting times. If we take α_1 as the initial condition (α^*), and α_{l+1} as a bad terminal state, then $\mathbb{P}(\mathcal{P}|\alpha_1) \leqslant \epsilon$, and hence

$$3M\epsilon^{1/l} \geqslant \min_{1\leqslant k\leqslant l} W_{\alpha_{k+1}\alpha_k}. \tag{B2}$$

Since $\epsilon < 1$, there is a nonzero probability of reaching a good terminal state and so Lemma 1 tells us there is a nonzero probability of reaching a bad terminal state. Taking such a path to a bad terminal state, the smallest rate along this path satisfies $0 < W_{\alpha_k,\alpha_{k+1}} \leqslant 3M\epsilon^{1/l}$. Without loss of generality, we can remove any loops in this path (places where $\alpha_r = \alpha_q$, $r \neq q$), leaving a loop-free path with nonzero probability, and since there are at most $(2N+2)^M$ states along this path we have $0 < W_{\alpha_k,\alpha_{k+1}} \leqslant 3M\epsilon^{(2N+2)^{-M}}$.

Now we progressively prune our rates. In particular, we formally set whatever parameter determines this smallest rate to zero. This parameter will either correspond to exactly one of the $f^{\pm}, k^{-}$. If it corresponds to a k^{+}, then all the g values involved must be smaller than $3M\epsilon^{(2N+2)^{-M}}$, and so set them all to zero in the new system. If it is possible to reach a good terminal state in this new system, it is possible to reach a bad terminal state by Lemma 1, and hence there exists a path with nonzero probability which, as before. So, we prune again. We repeat this procedure until there are no possible paths to the good terminal state (which must occur as there are only finitely many parameters). At this point, we can conclude that to reach the good terminal state, the system must make a transition where the rate is "slow." Typically that means that the transition rate is at most $3M\epsilon^{(2N+2)^{-M}}$, although in the case of a k^{+}, it could be the combination of $M-1$ small g values and is at most $3(M-1)M\epsilon^{(2N+2)^{-M}}$. The average time to reach a good terminal state, τ_g, satisfies $\tau \geqslant (1-\epsilon)\tau_g$, and hence if we can bound τ_g we can bound τ. To bound τ_g we can ask how long it takes on average to make one of these slow transitions, given that at least one of these transitions must be made to reach the good state. Since at most $3M$ transitions are possible from any given state (of which, at most M correspond to a $k^{\pm}$), even if the system was in a state where every possible transition was a slow transition, the rate at which a slow transition occurs would be at most $M^2(3M+3)\epsilon^{(2N+2)^{-M}}$. Thus, this quantity bounds the rate at which a slow transition occurs, whatever state the system is in. Hence, the average time for such a transition to occur is at least $M^{-2}(3M+3)^{-1}\epsilon^{-(2N+2)^{-M}}$, and hence in total, $\tau \geqslant M^{-2}(3M+3)^{-1}(1-\epsilon)\epsilon^{-(2N+2)^{-M}}$. Although not the tightest bound possible, it still shows that there must exist a speed-accuracy trade-off and $\tau \to \infty$ as $\epsilon \to 0$.

APPENDIX C: EXACT ASYMPTOTIC SYSTEM

Here we explicitly construct a solution that can achieve an arbitrarily small error rate. Motivated by the appearance of numerically optimized solutions, let us take $f^{+}(1,0) = \eta,\ f^{+}(i,s) = 1$ for $i > 1,\ f^{+}(1,1) = 0,\ f^{-}(i,0) = 0,\ \forall i,$ $f^{-}(1,1) = 1, f^{-}(i,1) = 0$ for $i > 1, k^{-} = 0, g(0) = g(1) = 0, g(i) = 1$ for $i > 1, \eta \ll 1$.

To compute the error rate, suppose that the first transition has occurred and hence a cell has transitioned from $(1,0)$ to $(2,0)$. At this point it will reach N, and contribute a constant signal $g = 1$ to its neighbors. Precisely when it transitions or whether it receives a signal is of no relevance for computing the error rate. Each of the next cells has a choice: they could transition to (2,0) with probability $\eta/(1+\eta)$ or transition to (1,1) with probability $1/(1+\eta)$. A transition to (2,0) guarantees the system will reach a bad terminal state, whereas a transition to (1,1) guarantees that cell will eventually reach (0,1) and will not signal to its neighbors. Thus, if it does transition to (0,1), in order for the system to reach the good terminal state the remaining cell has to also transition to (0,1), which occurs again with probability $1/(1+\eta)$. Hence, the probability of failure is $1 - 1/(1+\eta)^2 \approx 2\eta$.

The expected time to the first transition is $1/3\eta$; after this time, each cell can only ever make at most $N-1$ (for $N > 2$) transitions, each of which has a waiting time with mean at most 1. For fixed N this gives $\tau = 1/3\eta + O(1)$, or $\tau = 2/(3\epsilon) + O(1)$. This strategy is not asymptotically optimal, but it can be optimized by choosing $f^{\pm}(i,1)$ more carefully.

APPENDIX D: UNBOUNDED MUTUAL INFORMATION

In this section, we show that the full mutual information between a pair of trajectories in the system in Appendix C is unbounded. Let us consider $I(X_0^T; Y_0^T)$ with $X = u_1(t)$ and $Y = u_2(t)$. Using the chain rule for mutual information, we have that

$$\begin{aligned} I\big(X_0^T; Y_0^T, Z\big) &= I\big(X_0^T; Y_0^T\big) + I\big(X_0^T; Z \big| Y_0^T\big), \\ &= I\big(X_0^T; Z\big) + I\big(X_0^T; Y_0^T \big| Z\big), \end{aligned} \tag{D1}$$

for any random variable Z. Choosing a Z with a finite state space of size N_Z (and the size of this state space is independent of η) means that the entropy of Z, or entropy of Z conditioned on another variable, is bounded above by $H(Z) \leqslant \log N_Z$, and hence any mutual information term between Z and another variable is similarly bounded by $\log N_Z$. Hence, we can determine that, as $\epsilon \to 0$,

$$\begin{aligned} I\big(X_0^T; Y_0^T\big) \text{ bounded} &\iff I(X_0^T; Y_0^T, Z) \text{ bounded} \\ &\iff I\big(X_0^T, Z; Y_0^T, Z\big) \text{ bounded} \\ &\iff I\big(X_0^T; Y_0^T \big| Z\big) \text{ bounded.} \end{aligned} \tag{D2}$$

For the transition rates in Appendix C, the set of possible paths (sans waiting times) from the initial condition to the final is finite, so set Z to be the random variable representing which path is taken. Furthermore, let us take the path where u_1 first transitions from 1 to 2, followed by s_2 transitioning from 0 to 1, and then u_2 transitioning from 1 to 0. This particular path has an $O(1)$ probability of occurring, and so if the mutual information conditioned on this path diverges, then the overall mutual information diverges. Furthermore, from the data processing inequality, we can apply any coarse-graining function to the trajectories X_0^T and Y_0^T and only decrease the mutual information. Thus, if we reduce the trajectory X_0^T to the first time at which u_1 changes, and similarly Y_0^T to the first time at which u_2 changes, essentially we are left computing the mutual information, $I(\tau_1 : \hat{\tau})$, with $\hat{\tau} = \tau_1 + \tau_2 + \tau_3$, where all the τ_i's are drawn from exponential distributions, with rates 3η for τ_1, $(3+2\eta)$ for τ_2, and $(3+\eta)$ for τ_3. (Even after conditioning on a path the waiting time distribution is unchanged and reflects the sum of possible transition rates before conditioning.) Decomposing the mutual information,

$$\begin{aligned} I(\tau_1; \hat{\tau}) &= H(\hat{\tau}) - H(\hat{\tau}|\tau_1) \\ &\geqslant \max\{H(\tau_1), H(\tau_2), H(\tau_3)\} - H(\tau_2 + \tau_3) \\ &\geqslant H(\tau_1) - H(\tau_2) - H(\tau_3) \\ &= O(\log 1/\eta), \end{aligned} \tag{D3}$$

using the fact that $H(X+Y) \geqslant \max\{H(X), H(Y)\}$, $H(X,Y) \geqslant H(X+Y)$, $H(X+Y) = H(X) + H(Y)$ if X and Y are independent, and if $X \sim \text{Exp}(\lambda)$ then $H(X) = 1 - \log\lambda$. Thus, the mutual information diverges at least as fast as $O(\log 1/\eta)$.

Intuitively, in this system cell 1 can transition at any point in an $O(1/\eta)$ time. If you observe a transition $1 \to 0$ in cell

2 at time T, you know that with $O(1)$ probability, cell 1 transitioned at a time $T + O(1)$, narrowing down from an $O(1/\eta)$ uncertainty and thus gaining a significant amount of information. However, this intuition also suggests a strategy that would have a finite amount of mutual information, while still having an arbitrarily small error rate. If we keep $g(i) = 1$ for $i > 1$ but set $f^-(1, 1) = \eta$, $f^+(i, s) = \eta$ for $i > 1$, the error rate is unaffected. However, observing a transition $1 \to 0$ in cell 2 does not narrow down the time at which another cell transitioned all that much, since the cell 2 transition occurs at a time $O(1/\eta)$ after the first transition. If we take the information, $I((u_1, s_1); (u_2, s_2))$, this would again diverge. In some sense, in this new system there is divergent mutual information; it is just not stored in the internal state. We suspect that the mutual information $I((u_1, s_1); (u_2, s_2))$ always becomes divergent as the error goes to zero.

APPENDIX E: CONDITIONAL DYNAMICAL INFORMATION BETWEEN LATERALLY INHIBITED STATES

In Sec. IV, we showed that successful patterning depends on asymmetric, nontrivial communication among cells. While Fig. 4 depicts transfer between inhibitor and inhibited cells, Fig. 9 illustrates communication among the inhibited cells themselves. Unlike the asymmetric flow involving inhibitor cells, the exchange here is symmetric, reflecting the equivalence of inhibited cells.

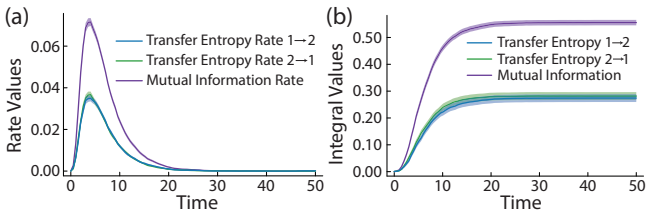

FIG. 9. Conditioning on final cell fates reveals nontrivial communication between the laterally inhibited states. (a) Conditioned mutual information and transfer entropy rates between the inhibited cell 1 and inhibited cell 2 as well as (b) the corresponding integrals. Throughout, each computation is averaged over $n = 10\,000$ Monte Carlo samples, and the shaded regions show $1.96\times$standard error.

[1] D. B. Brückner and G. Tkačik, Information content and optimization of self-organized developmental systems, Proc. Natl. Acad. Sci. USA **121**, e2322326121 (2024).

[2] F. Schweisguth and F. Corson, Self-organization in pattern formation, Dev. Cell **49**, 659 (2019).

[3] A. Kicheva and J. Briscoe, Control of tissue development by morphogens, Annu. Rev. Cell Dev. Biol. **39**, 91 (2023).

[4] M. D. Petkova, G. Tkačik, W. Bialek, E. F. Wieschaus, and T. Gregor, Optimal decoding of cellular identities in a genetic network, Cell **176**, 844 (2019).

[5] I. Hajji, F. Corson, and W. Keil, Quantitative guiding of developmental cell fate transitions using a dynamical landscape model, bioRxiv, doi:10.1101/2024.11.16.623539.

[6] W. Wang, K. Ni, D. Poe, and J. Xing, Transiently increased coordination in gene regulation during cell phenotypic transitions, PRX Life **2**, 043009 (2024).

[7] T. A. Leathers and C. D. Rogers, Time to go: Neural crest cell epithelial-to-mesenchymal transition, Development **149**, dev200712 (2022).

[8] S. J. Bray and A. Bigas, Modes of notch signalling in development and disease, Nat. Rev. Mol. Cell Biol. **26** 522 (2025).

[9] M.-S. Phan, J.-m. Kim, C. Picciotto, L. Couturier, N. Veits, K. Mazouni, and F. Schweisguth, Symmetry breaking and fate divergence during lateral inhibition in *drosophila*, Development **151**, dev203165 (2024).

[10] F. Corson, L. Couturier, H. Rouault, K. Mazouni, and F. Schweisguth, Self-organized notch dynamics generate stereotyped sensory organ patterns in *Drosophila*, Science **356**, eaai7407 (2017).

[11] F. Bocci, J. N. Onuchic, and M. K. Jolly, Understanding the principles of pattern formation driven by notch signaling by integrating experiments and theoretical models, Front. Physiol. **11**, 929 (2020).

[12] E. D. Siggia and M. Vergassola, Decisions on the fly in cellular sensory systems, Proc. Natl. Acad. Sci. USA **110**, E3704 (2013).

[13] T. J. Kobayashi, Implementation of dynamic bayesian decision making by intracellular kinetics, Phys. Rev. Lett. **104**, 228104 (2010).

[14] K. Nakamura and T. J. Kobayashi, Connection between the bacterial chemotactic network and optimal filtering, Phys. Rev. Lett. **126**, 128102 (2021).

[15] T. Tottori and T. J. Kobayashi, Theory for optimal estimation and control under resource limitations and its applications to biological information processing and decision-making, Phys. Rev. Res. **7**, 043048 (2025).

[16] F. Corson and E. D. Siggia, Geometry, epistasis, and developmental patterning, Proc. Natl. Acad. Sci. USA **109**, 5568 (2012).

[17] P. W. Sternberg, Vulval development, *WormBook*, ed. The *C. elegans* Research Community, WormBook, 2005, doi:10.1895/wormbook.1.6.1.

[18] A. Roure, P. Lemaire, and S. Darras, An otx/nodal regulatory signature for posterior neural development in ascidians, PLoS Genet. **10**, e1004548 (2014).

[19] R. Bettoni, G. Dupont, A. M. Walczak, and S. de Buyl, Optimizing information transmission in neural induction constrains cell surface contacts of ascidian embryos, PRX Life **3**, 033004 (2025).

[20] M. Kato and T. J. Kobayashi, Optimality theory of stigmergic collective information processing by chemotactic cells, PRX Life **3**, 043005 (2025).

[21] N. Ohta and Y. Satou, Multiple signaling pathways coordinate to induce a threshold response in a chordate embryo, PLoS Genet. **9**, e1003818 (2013).

[22] T. Tottori and T. J. Kobayashi, Memory-limited partially observable stochastic control and its mean-field control approach, Entropy **24**, 1599 (2022).
[23] T. Rashid, M. Samvelyan, C. S. De Witt, G. Farquhar, J. Foerster, and S. Whiteson, Monotonic value function factorisation for deep multi-agent reinforcement learning, J. Mach. Learn. Res. **21**, 1 (2020).
[24] M. Samvelyan, T. Rashid, C. S. de Witt, G. Farquhar, N. Nardelli, T. G. J. Rudner, C.-M. Hung, P. H. S. Torr, J. Foerster, and S. Whiteson, The starcraft multi-agent challenge, arXiv:1902.04043.
[25] E. H. Durfee, Distributed problem solving and planning, in *Multi-Agent Systems and Applications* (Springer, Berlin, 2001), pp. 118–149.
[26] J. N. Gray, Notes on data base operating systems, in *Operating Systems: An Advanced Course* (Springer, Berlin, 1978), pp. 393–481.
[27] S. Strogatz, S. Walker, J. M. Yeomans, C. Tarnita, E. Arcaute, M. De Domenico, O. Artime, and K.-I. Goh, Fifty years of 'more is different', Nat. Rev. Phys. **4**, 508 (2022).
[28] P. W. Anderson, More is different, Science **177**, 393 (1972).
[29] J. R. Collier, N. A. Monk, P. K. Maini, and J. H. Lewis, Pattern formation by lateral inhibition with feedback: A mathematical model of delta-notch intercellular signalling, J. Theor. Biol. **183**, 429 (1996).
[30] K. Friston, M. Levin, B. Sengupta, and G. Pezzulo, Knowing one's place: A free-energy approach to pattern regulation, J. R. Soc. Interface **12**, 20141383 (2015).
[31] R. Kuintzle, L. A. Santat, and M. B. Elowitz, Diversity in notch ligand–receptor signaling interactions, eLife **12**, RP91422 (2024).
[32] A.-L. Moor and C. Zechner, Dynamic information transfer in stochastic biochemical networks, Phys. Rev. Res. **5**, 013032 (2023).
[33] H. K. Hartline, H. G. Wagner, and F. Ratliff, Inhibition in the eye of limulus, J. Gen. Physiol. **39**, 651 (1956).
[34] O. Barad, D. Rosin, E. Hornstein, and N. Barkai, Error minimization in lateral inhibition circuits, Sci. Signaling **3**, ra51 (2010).
[35] D. Riddle, T. Blumenthal, B. Meyer, and J. Priess, *C. Elegans II* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1997).
[36] P. François and E. D. Siggia, A case study of evolutionary computation of biochemical adaptation, Phys. Biol. **5**, 026009 (2008).
[37] B. Zoller, A. Bénichou, T. Gregor, and G. Tkačik, Invariant non-equilibrium dynamics of transcriptional regulation optimize information flow, arXiv:2507.12395.
[38] A. Hilfinger, T. M. Norman, G. Vinnicombe, and J. Paulsson, Constraints on fluctuations in sparsely characterized biological systems, Phys. Rev. Lett. **116**, 058101 (2016).
[39] B. Kell, R. Ripsman, and A. Hilfinger, Noise properties of adaptation-conferring biochemical control modules, Proc. Natl. Acad. Sci. USA **120**, e2302016120 (2023).
[40] J. Yan, A. Hilfinger, G. Vinnicombe, and J. Paulsson, Kinetic uncertainty relations for the control of stochastic reaction networks, Phys. Rev. Lett. **123**, 108101 (2019).
[41] D. Zhang, Y. Cao, Q. Ouyang, and Y. Tu, The energy cost and optimal design for synchronization of coupled molecular oscillators, Nat. Phys. **16**, 95 (2020).
[42] D. Zhang, Y. Cao, Q. Ouyang, and Y. Tu, Altruistic resource-sharing mechanism for synchronization: The energy-speed-accuracy trade-off, Phys. Rev. Lett. **135**, 037401 (2025).
[43] See Supplemental Material at http://link.aps.org/supplemental/10.1103/lfnt-8qbt for additional algorithmic and numerical details, Figs. S1–S10, and Tables S1–S2, which includes Refs. [82–89].
[44] Z.-q. Luo, W.-k. Ma, A. M.-c. So, Y. Ye, and S. Zhang, Semidefinite relaxation of quadratic optimization problems, IEEE Signal Process. Mag. **27**, 20 (2010).
[45] J. A. Owen, T. R. Gingrich, and J. M. Horowitz, Universal thermodynamic bounds on nonequilibrium response with biochemical applications, Phys. Rev. X **10**, 011066 (2020).
[46] M. Razo-Mejia, S. Marzen, G. Chure, R. Taubman, M. Morrison, and R. Phillips, First-principles prediction of the information processing capacity of a simple genetic circuit, Phys. Rev. E **102**, 022404 (2020).
[47] J. O. Dubuis, G. Tkačik, E. F. Wieschaus, T. Gregor, and W. Bialek, Positional information, in bits, Proc. Natl. Acad. Sci. USA **110**, 16301 (2013).
[48] G. Tkačik, A. M. Walczak, and W. Bialek, Optimizing information flow in small genetic networks, Phys. Rev. E **80**, 031920 (2009).
[49] A. M. Walczak, G. Tkačik, and W. Bialek, Optimizing information flow in small genetic networks. II. Feed-forward interactions, Phys. Rev. E **81**, 041905 (2010).
[50] G. Tkačik, A. M. Walczak, and W. Bialek, Optimizing information flow in small genetic networks. III. A self-interacting gene, Phys. Rev. E **85**, 041903 (2012).
[51] L. Hahn, A. M. Walczak, and T. Mora, Dynamical information synergy in biochemical signaling networks, Phys. Rev. Lett. **131**, 128401 (2023).
[52] G. Tkačik and P. R. ten Wolde, Information processing in biochemical networks, Annu. Rev. Biophys. **54**, 249 (2025).
[53] Y. Tang, A. Adelaja, F. X. F. Ye, E. Deeds, R. Wollman, and A. Hoffmann, Quantifying information accumulation encoded in the dynamics of biochemical signaling, Nat. Commun. **12**, 1272 (2021).
[54] O. Witteveen, S. J. Rosen, R. S. Lach, M. Z. Wilson, and M. Bauer, Optimizing information transmission in the canonical wnt pathway, arXiv:2506.22633.
[55] M. Kramar, L. Hahn, A. M. Walczak, T. Mora, and M. Coppey, Single cells can resolve graded stimuli, PRX Life **3**, 043016 (2025).
[56] R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, and A. Levchenko, Information transduction capacity of noisy biochemical signaling networks, Science **334**, 354 (2011).
[57] M. Bauer, M. D. Petkova, T. Gregor, E. F. Wieschaus, and W. Bialek, Trading bits in the readout from a genetic network, Proc. Natl. Acad. Sci. USA **118**, e2109011118 (2021).
[58] T. R. Sokolowski, T. Gregor, W. Bialek, and G. Tkačik, Deriving a genetic regulatory network from an optimization principle, Proc. Natl. Acad. Sci. USA **122**, e2402925121 (2025).
[59] A.-L. Moor, A. Tjalma, M. Reinhardt, P. Rein ten Wolde, and C. Zechner, State- versus reaction-based information processing in biochemical networks, arXiv:2505.13373.
[60] F. Tostevin and P. Rein ten Wolde, Mutual information between input and output trajectories of biochemical networks, Phys. Rev. Lett. **102**, 218101 (2009).

[61] A. Das and P. Rein ten Wolde, Exact computation of transfer entropy with path weight sampling, Phys. Rev. Lett. **135**, 107404 (2025).

[62] M. Reinhardt, G. Tkačik, and P. Rein ten Wolde, Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories, Phys. Rev. X **13**, 041017 (2023).

[63] L. Duso and C. Zechner, Selected-node stochastic simulation algorithm, J. Chem. Phys. **148**, 164108 (2018).

[64] J. R. Norris, Continuous-time Markov chains II, in *Markov Chains*, Cambridge Series in Statistical and Probabilistic Mathematics (Cambridge University Press, Cambridge, UK, 1997), pp. 108–127.

[65] S. Watanabe, Information theoretical analysis of multivariate correlation, IBM J. Res. Dev. **4**, 66 (1960).

[66] H. H. Mattingly, K. Kamino, B. B. Machta, and T. Emonet, *Escherichia coli* chemotaxis is information limited, Nat. Phys. **17**, 1426 (2021).

[67] H. B. Brandl, J. A. Klarevas-Irby, D. Zuñiga, C. Hansen Wheat, C. Christensen, F. Omengo, C. Nzomo, W. Cherono, B. Nyaguthii, and D. R. Farine, The physiological cost of leadership in collective movements, Curr. Biol. **35**, 4003 (2025).

[68] A. Kraskov, H. Stögbauer, and P. Grassberger, Estimating mutual information, Phys. Rev. E **69**, 066138 (2004).

[69] C. Schwayer, S. Barbiero, D. B. Brückner, C. Baader, N. A. Repina, O. E. Diaz, L. C. Meylan, V. Kalck, S. Suppinger, Q. Yang, J. Schnabl, U. Kilik, J. G. Camp, B. Stockinger, M. Bühler, M. B. Stadler, E. Hannezo, and P. Liberali, Cell heterogeneity and fate bistability drive tissue patterning during intestinal regeneration, bioRxiv, doi:10.1101/2025.01.14.632683.

[70] S. Okuda, T. Miura, Y. Inoue, T. Adachi, and M. Eiraku, Combining turing and 3D vertex models reproduces autonomous multicellular morphogenesis with undulation, tubulation, and branching, Sci. Rep. **8**, 2386 (2018).

[71] C. M. Glen, M. L. Kemp, and E. O. Voit, Agent-based modeling of morphogenetic systems: Advantages and challenges, PLoS Comput. Biol. **15**, e1006577 (2019).

[72] S. Bajpai, R. Chelakkot, R. Prabhakar, and M. M. Inamdar, Role of delta-notch signalling molecules on cell–cell adhesion in determining heterogeneous chemical and cell morphological patterning, Soft Matter **18**, 3505 (2022).

[73] T. Dullweber and A. Erzberger, Mechanochemical feedback loops in contact-dependent fate patterning, Curr. Opin. Syst. Biol. **32-33**, 100445 (2023).

[74] M. Smart and A. Zilman, Emergent properties of collective gene-expression patterns in multicellular systems, Cell Rep. Phys. Sci. **4**, 101247 (2023).

[75] M. Faldor and A. Cully, CAX: Cellular automata accelerated in JAX, in *International Conference on Learning Representations* (OpenReview.net, Singapore, 2025).

[76] A. D. Richardson, T. Antal, R. A. Blythe, and L. J. Schumacher, Learning spatio-temporal patterns with neural cellular automata, PLOS Comput. Biol. **20**, e1011589 (2024).

[77] M. Reinhardt, G. Tkačik, and P. Rein ten Wolde, Ml-pws: Estimating the mutual information between experimental time series using neural networks, arXiv:2508.16509.

[78] N. C. Lammers, Y. J. Kim, J. Zhao, and H. G. Garcia, A matter of time: Using dynamics and theory to uncover mechanisms of transcriptional bursting, Curr. Opin. Cell Biol. **67**, 147 (2020).

[79] E. Dessaud, L. L. Yang, K. Hill, B. Cox, F. Ulloa, A. Ribeiro, A. Mynett, B. G. Novitch, and J. Briscoe, Interpretation of the sonic hedgehog morphogen gradient by a temporal adaptation mechanism, Nature (London) **450**, 717 (2007).

[80] C. Marshall, Specificity of receptor tyrosine kinase signaling: Transient versus sustained extracellular signal-regulated kinase activation, Cell **80**, 179 (1995).

[81] A. Tripathi, J. Dunkel, and D. Skinner, Collective is different: Information exchange and speed-accuracy tradeoffs in self-organized patterning [data set], Zenodo (2026), doi:10.5281/zenodo.18189941.

[82] J. Bragantini, I. Theodoro, X. Zhao, T. A. P. M. Huijben, E. Hirata-Miyasaki, S. VijayKumar, A. Balasubramanian, T. Lao, R. Agrawal, S. Xiao, J. Lammerding, S. Mehta, A. X. Falcão, A. Jacobo, M. Lange, and L. A. Royer, Ultrack: Pushing the limits of cell tracking across biological scales, Nat. Methods **22**, 2423 (2025).

[83] B. Efron and C. Stein, The jackknife estimate of variance, Ann. Statist. **9**, 586 (1981).

[84] I. Loshchilov and F. Hutter, SGDR: Stochastic gradient descent with warm restarts, in *International Conference on Learning Representations* (OpenReview.net, Toulon, France. 2017).

[85] S. Mohamed, M. Rosca, M. Figurnov, and A. Mnih, Monte Carlo gradient estimation in machine learning, arXiv:1906.10652.

[86] J. Nocedal and S. Wright, *Numerical optimization*, Springer Series in Operations Research and Financial Engineering (Springer, New York, 2006).

[87] S. J. Reddi, S. Kale, and S. Kumar, On the convergence of Adam and beyond, *International Conference on Learning Representations* (OpenReview.net, Vancouver, BC, Canada, 2018).

[88] R. E. Spinney, J. T. Lizier, and M. Prokopenko, Transfer entropy in physical systems and the arrow of time, Phys. Rev. E **94**, 022135 (2016).

[89] R. E. Spinney, M. Prokopenko, and J. T. Lizier, Transfer entropy in continuous time, with applications to jump and neural spiking processes, Phys. Rev. E **95**, 032319 (2017).