Title

Minimal Network Topologies for Signal Processing during Collective Cell Chemotaxis.

Journal

Biophysical Journal, 114(12)

Authors

Yue, Haicen

Camley, Brian

Rappel, Wouter-Jan

Publication Date

2018-06-19

DOI

10.1016/j.bpj.2018.04.020

Peer reviewed

Biophysical *Journal*
Article

Biophysical Society

Minimal Network Topologies for Signal Processing during Collective Cell Chemotaxis

Haicen Yue,[1] Brian A. Camley,[2,3] and Wouter-Jan Rappel[1,*]
[1]Department of Physics, University of California, San Diego, La Jolla, California; [2]Department of Physics and Astronomy and [3]Department of Biophysics, Johns Hopkins University, Baltimore, Maryland

ABSTRACT Cell-cell communication plays an important role in collective cell migration. However, it remains unclear how cells in a group cooperatively process external signals to determine the group's direction of motion. Although the topology of signaling pathways is vitally important in single-cell chemotaxis, the signaling topology for collective chemotaxis has not been systematically studied. Here, we combine mathematical analysis and simulations to find minimal network topologies for multicellular signal processing in collective chemotaxis. We focus on border cell cluster chemotaxis in the *Drosophila* egg chamber, in which responses to several experimental perturbations of the signaling network are known. Our minimal signaling network includes only four elements: a chemoattractant, the protein Rac (indicating cell activation), cell protrusion, and a hypothesized global factor responsible for cell-cell interaction. Experimental data on cell protrusion statistics allows us to systematically narrow the number of possible topologies from more than 40,000,000 to only six minimal topologies with six interactions between the four elements. This analysis does not require a specific functional form of the interactions, and only qualitative features are needed; it is thus robust to many modeling choices. Simulations of a stochastic biochemical model of border cell chemotaxis show that the qualitative selection procedure accurately determines which topologies are consistent with the experiment. We fit our model for all six proposed topologies; each produces results that are consistent with all experimentally available data. Finally, we suggest experiments to further discriminate possible pathway topologies.

INTRODUCTION

Collective cell migration plays an important role in many biological processes, including development, wound healing, and cancer metastasis, and it has been the focus of much recent experimental and theoretical work (1–7). Interestingly, collective migration is not merely the result of many independent cells moving around but can exhibit unique behavior. For example, measurements of several different cell types have shown that cells cooperate to sense a gradient such that cell clusters follow gradients whereas single cells are unable to detect this gradient (8,9). Furthermore, expanding cell monolayers often reveal the spontaneous formation of finger-like instabilities with specialized leader cells at their tips (10,11). These results point toward the important role that cell-cell interactions can play in collective migration. Here, we are particularly interested in collective chemotaxis, or how clusters of cells work together to follow a chemical gradient, as measured in malignant lymphocytes (8), neural crest migration in developing embryos (9), and border cell migration in *Drosophila* (12–14).

Collective chemotaxis problems have been modeled by many groups, with widely varying assumptions for cell-cell interactions. These include a focus on intercellular forces such as cell-cell repulsion and adhesion (15), mechanical communication via contact inhibition of locomotion (16–19), or biochemical communication through a diffusible inhibitor (20,21). In particular, several earlier theoretical studies show that, depending on the biochemical details of how the cluster processes the signal, cluster speeds and directionalities can be profoundly different (20–23). Current experiments have not yet provided a way to clearly determine these signal-processing mechanisms, and most studies have assumed only a minimal multicellular local excitation global inhibition (LEGI) signal-processing mechanism (20,21,23).

In this article, we develop a method to use existing experimental data to constrain possible multicellular signaling networks. Our approach is a qualitative perturbation analysis in which we use mathematical analysis to determine how experimental interventions such as photoactivation of one element of the signaling network will affect measured

Submitted January 12, 2018, and accepted for publication April 10, 2018.
*Correspondence: rappel@physics.ucsd.edu
Editor: Cynthia Reinhart-King.
https://doi.org/10.1016/j.bpj.2018.04.020

outcomes. These behaviors will often only depend on the topology of the signaling network—which elements are connected to which—and not the biochemical details. We use this technique to determine potential minimal network topologies for border cell cluster migration in the *Drosophila* egg chamber. We can show that existing experimental data rule out all networks with five or fewer interactions (including the aforementioned LEGI model), and we can deduce that only six possible six-link topologies are able to explain the experimental data. In this process, we introduce two methods to circumvent the complexity caused by a stochastic, spatially extended problem like border cell chemotaxis. We first use analytical methods for the selection of possible topologies to avoid exhaustively simulating all topologies. Then, for selected topologies, we carry out quantitative simulations in which we solve nonlinear stochastic differential equations using a master equation approach instead of direct numerical integration. We verify through explicit numerical sampling that these networks are fully consistent with the data and carry out a parameter fit for the six topologies. Finally, we propose additional experiments to further narrow down the possible options.

METHODS

The ordinary differential equation (ODE) for Rac is:

$$\frac{d[R_i]}{dt} = \frac{\left(basal_{AR} + k_{SR}[S_i] + k_{GR}[G]\right)}{1 + PI_i}\left(R^{tot} - [R_i]\right) - \left(basal_{DR} + k_{-GR}[G]\right)[R_i].$$

Here, $[R_i]$ represents the concentration of Rac in cell i, and $[S_i]$ is the signal sensed by cell i. No matter what the signal profile across the whole egg chamber looks like, as the size of the cluster is small compared to the length of the egg chamber, it is reasonable to approximate the $[S_i]$ within the cluster in a linear form as follows: $[S_i] = S_{mean}(1 + S_{grad}\cos\theta_i)$, with θ_i as the angle of cell i in the cluster. In this equation, the first term is the activation of Rac, and the second term is the deactivation of Rac. $basal_{AR}$ and $basal_{DR}$ are the basal activation and deactivation rates, respectively. The activation or deactivation terms with $[S_i]$ or $[G]$ are the rates regulated by the signal or the global factor. PI_i is the photoinhibition effect implemented on Rac in cell i. The photoinhibition works with a photoactivatable form of the dominant-negative mutants of Rac (PA-RacT17N). RacT17N strongly binds to upstream guanine nucleotide exchange factors and blocks wild-type Rac from being activated. We thus divide the activation term in this equation by $(1 + PI_i)$ to describe the photoinhibition effect.

The ODE for the global factor G is:

$$\frac{d[G]}{dt} = \left(basal_{AG} + k_{RG}\left([R]_{avg} + PA_{avg}\right) + k_{SG}[S]_{avg} + k_{PG} \times NP\right) - \left(basal_{DG} + k_{-RG}\left([R]_{avg} + PA_{avg}\right) + k_{-SG}[S]_{avg} + k_{-PG} \times NP\right)[G].$$

Here, $[R]_{avg}$ and $[S]_{avg}$ are the average concentrations of Rac and signal over all the cells. In this equation, the first term is the production of the global factor, and the second term is the degradation of this factor. $basal_{AG}$ and $basal_{DG}$ are the basal rates of G. PA_i is the photoactivation effect on Rac in cell i. In the photoactivatable form of Rac, the binding site of downstream effector on Rac is blocked without light of certain wavelength, and light-induced conformational change would release this block and make Rac available to the effectors (24). This means light treatment provides more activated Rac available to induce downstream event. This photoactivation effect does not go into the equation of $[R]$. Instead, we add the additional activated Rac (PA) provided by the light treatment to wherever there is $[R]$ in the equations of $[G]$ and $[P]$ in the form of $[R] + PA$. PA_{avg} is the average effect of photoactivation and is defined as $PA_{avg} = (1/n)\sum_i PA_i$, where n is the number of cells in a cluster. In addition, we have not introduced G^{tot} in the $[G]$ equations, as we assume that $[G]$ is far from saturation.

The stochastic differential equation (SDE) for protrusion is:

$$\frac{d[P_i]}{dt} = \left(k_{RP}\left([R_i] + PA_i\right) + k_{GP}[G] + k_{SP}[S_i]\right)\frac{[P_i]^2}{[P_i]^2 + 1^2}\left(P^{tot} - [P_i]\right) - \left(basal_{DP} + k_{-GP}[G] + k_{-RP}\left([R_i] + PA_i\right) + k_{-SP}[S_i]\right)[P_i] + \eta(t).$$

Here, the first term is the activation of P, and the second-order Hill function represents the self-activation of P. The second term is the deactivation of P, and the basal deactivation rate $basal_{DP}$ is normalized to 1 in the simulation. The third term is the Gaussian noise term.

If we assume that the equations for $[R]$ and $[G]$ reach equilibrium quickly compared to the timescale of protruding (20–30 min) (13,25), we can get the steady states for $[R]$ and $[G]$ as follows and then plug them into the $[P]$ equations:

$$[R_i]^s = \frac{basal_{AR} + [S_i] + k_{GR}[G]^s}{basal_{AR} + [S_i] + k_{GR}[G]^s + \left(basal_{DR} + k_{-GR}[G]^s\right)\left(1 + PI_i\right)}R^{tot}$$

$$[G]^s = \frac{k_{RG}\left([R]^s_{avg} + PA_i\right) + k_{SG}[S]_{avg} + k_{PG} \times NP + basal_{AG}}{k_{-RG}\left([R]^s_{avg} + PA_i\right) + k_{-SG}[S]_{avg} + k_{-PG} \times NP + basal_{DG}}.$$

With this fast equilibrium assumption, we are only interested in the steady states of [R] and [G] so that we can normalize $basal_{DR}$, R^{tot}, $basal_{AG}$, and $basal_{DG}$ to 1.

It is worth noting that these equations are the general forms, including all the possible interactions, and given a certain network, only the parameters related to the existing interactions are nonzero. For networks with five or six interactions, three of the interactions are fixed, and their related parameters are listed in Table 2. The rest of the parameters related to the undetermined interactions are represented by k_1, k_2, and k_3 in Table 2. For one specific network, such as network A in Fig. 2 *c*, $k_1 \equiv k_{SG}$, $k_2 \equiv k_{RG}$, and $k_3 \equiv k_{-GP}$.

RESULTS

Border cell migration: available experimental data

We study a characteristic example of collective chemotaxis: the guided movement of border cell clusters during the embryonic stage of *Drosophila* (26–28). A border cell cluster consists of two polar cells and 4–8 border cells (Fig. 1 *a*) and migrates during development between the nurse cells toward the oocyte as an interconnected group. During this migration, only the border cells show protrusions, whereas the polar cells do not protrude and remain at the center of the cluster. The directional migration of the cluster is essential for *Drosophila* oogenesis and is guided by chemoattractant gradients (29,30). Using genetic manipulations of the signal receptors and photoactivatable Rac, a Rho GTPase known to be pivotal in the control and formation of protrusions during cell migration (31), experiments show that the change of signaling state in one cell of the cluster can guide the movement of the whole cluster (32). Light-mediated activation and inhibition of Rac in a single cell can not only change the direction of the cluster but also inhibit or promote the protrusion of other cells in the cluster (13).

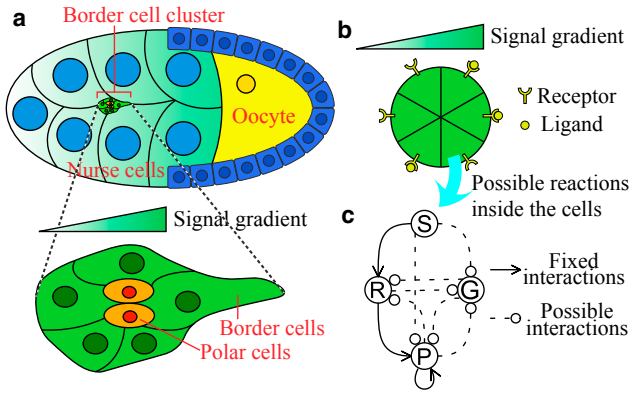

FIGURE 1 Border cell cluster in *Drosophila* egg chamber and model simplification. (*a*) A schematic drawing shows a *Drosophila* egg chamber and a border cell cluster composed of border cells and polar cells. The cluster moves toward the oocyte guided by chemical signals. (*b*) Our simplified model of a border cell cluster consists of six nondeformable cells arranged in a rigid array. (*c*) The basic structure of the signal-processing network consists of four nodes, three fixed positive interactions (*solid arrows*), and seven possible interactions (*dashed arrows* with round heads, meaning that the sign of the interactions is undetermined). To see this figure in color, go online.

We use experimental data obtained from (13), which is summarized in Table 1. These data quantify the number of border cells protruding and the directionality index of the cluster, which can be either positive, corresponding to more protruding cells at the front of the cluster, or negative, corresponding to more protruding cells located away from the highest chemoattractant concentration (13) (a detailed definition of the directionality index will be given later). In wild-type clusters, the average number of cells containing a protrusion is about two, whereas the directionality index is large and positive. The authors of (13) also obtained additional data using transgenic flies in which Rac can be selectively and locally activated or inhibited using light. Photoinhibition of Rac of a border cell at the front of the cluster, for example, results in a larger number of average protrusions in the noninhibited cells and a directionality close to zero. In addition, the number of protrusions was quantified in cells expressing dominant negative forms of the chemoattractant receptors, denoted here by $Receptor^{DN}$. In these cells, ligands can still bind to receptors but fail to deliver a downstream signal. Clusters expressing $Receptor^{DN}$ were found to have more protrusions on average than wild-type clusters but with a roughly zero directionality index, as would be expected in the absence of any directional signal.

Model

To construct our model, because the experimental measurements we study are the protrusions in the cluster measured over a relatively short time, we will not consider cell deformation, cluster motion, or cluster rearrangement. Furthermore, because polar cells do not protrude and remain at the center of the cluster, we will only consider border cells in our simulations. We fix the number of these cells to six but note that the analytical methods do not rely on the number of cells in a cluster, and we have verified that the numerical simulation results remain similar for clusters with more or fewer cells. The resulting model is schematically shown in Fig. 1 *b* and consists of a rigid cluster with cells making an angle of $\theta_i = (\pi/3)i$, i = 1,..,6 with the gradient direction. Each border cell i contains an identical pathway, and our goal is to determine a minimal set of components or nodes required to capture experimental findings. Because the cell is responding to an external chemoattractant in the form of protrusions and because the activation and inhibition of Rac affect the cell's response, we assume that the pathway has as local components the input chemoattractant signal S_i, the protrusion level P_i, and the Rac activation level R_i. In addition, we introduce a fourth global component, G, representing cell-cell communication. We do not specify the nature of this communication but note that it can be in the form of a small molecule, as suggested by experiments on branching morphogenesis (20) or in the form of mechanical interactions (12,33,34). Because in either case

TABLE 1 Summary of Experimental Data, Uncertainties Indicate SD

Experiment (Shorthand)	Number of Protruding Cells	Directionality Index	Comments
Wild-type (WT)	2.0 ± 0.2	0.435 ± 0.035	
Photoinhibition of WT (inhibit)	4.5 ± 0.6	-0.025 ± 0.095	photoinhibition of Rac in the cell at the front of the wild-type cluster
$\text{Receptor}^{\text{DN}}$ (RecDN)	4.15 ± 0.35	0.015 ± 0.085	cells expressing the dominant negative form of the chemoattractant receptors
Photoactivation of WT ($\text{activate}_{\text{back}}$)	2.0 ± 0.3	-0.405 ± 0.055	photoactivation of Rac in the cell at the back of the wild-type cluster
Photoactivation of $\text{Receptor}^{\text{DN}}$ ($\text{activate}_{\text{RecDN}}$)	2.1 ± 0.4	0.245 ± 0.045	photoactivation of Rac in one cell of the $\text{Receptor}^{\text{DN}}$ cluster

The SD was extracted from the figures in (13).

communication between cells is fast compared to protrusion dynamics, which occur on timescales of 20–30 min (13,25), we will assume that all cells in the cluster share the same value of G. Note that this assumption is valid as long as the degradation of G is not too fast (21,22). The output of the model and the quantities that can be compared to experimental data include the total number of cells with protrusions, represented by NP, and the direction of the protrusions, quantified by a directionality index specified in Further Discrimination between the Six Minimal Networks with Six Interactions.

We will consider three possible interactions or links between each node of the network, including self-interactions: 1) positive, in which the output of one node increases the activity of the other node or of itself; 2) negative, corresponding to a repression of the activity; and 3) no interaction. As a result, we have to consider 3^{16} possible topologies for the model network. One potential technique to determine if these networks are consistent with the data would be to follow the approach of (35), in which the researchers exhaustively searched three-node networks, sampling a broad range of parameters for each to find those with perfect adaptation to a changing signal. However, this is not computationally feasible given the more than 40,000,000 possibilities at hand. Such an approach is further complicated by the fact that our model describes a stochastic, spatially extended system. Thus, although we describe the pathway as a four-node network in which one node is the (fixed) chemotactic input, the equations for the entire cluster contain 13 variables and 13 equations (two local equations for each cell and one global equation). Instead of the sampling approach of (35), we will pursue a qualitative perturbation analysis method.

Qualitative selection approach

We will now reduce the number of possible topologies based on experimental data and without specifying specific functional forms of the interactions. In this qualitative selection procedure, we do not take into account the precise number of protrusions after a specific experimental perturbation. Instead, we only take into account if the perturbation increases or decreases the number of protrusions along with the location of the protrusions. The flow of this procedure is schematically shown in Fig. 2 *a*. As we will see below, we only have to consider topologies with fewer than seven links. Therefore, we start with all possible network configurations with five-link (amounting to 139,776) and six-link (totaling 512,512) topologies.

Before addressing the detailed results in Wang et al., we can already rule out a large amount of these topologies. First, we note the general result that Rac is known to be activated by an external signal and activates downstream effectors of cell protrusion (31). Consequently, positive interactions from S to R and from R to P are required. Second, experiments have shown that positive feedback loops that are downstream of activated Rac help maintain the protrusion (31). These loops will be captured by a positive self-activation link for P. In addition, as there is no experimental evidence for the amplification of R or G and as we are trying to find the minimal topologies, we assume that there is no self-activation or self-inhibition for these nodes. Furthermore, because S represents the external signal, we assume that there is no feedback to S. This reduction results in the networks that are shown in Fig. 1 *c*, where the three fixed positive interactions are shown as black solid arrows and where the seven remaining possible interactions are shown using black dotted arrows with round heads, meaning that the sign of the interactions is undetermined. Finally, we note that the experiments of Wang et al. show that local photoactivation and inhibition affect the protrusion number of the entire cluster. This automatically implies that there must be cell-cell communication. In our model, this communication is achieved through the global factor G, and we thus need at least two more links in addition to the three shown in Fig. 1 *c*: one input to G and one output from G. Taking into account the restrictions mentioned above, we find that we have 24 possible five-link topologies and 168 possible six-link topologies, schematically indicated by the second step in Fig. 2 *a*. Of course, there will be even more topologies with a larger number of links. However, as we will see below, several six-link topologies are consistent with the experimental data, and because we are searching for minimally complex

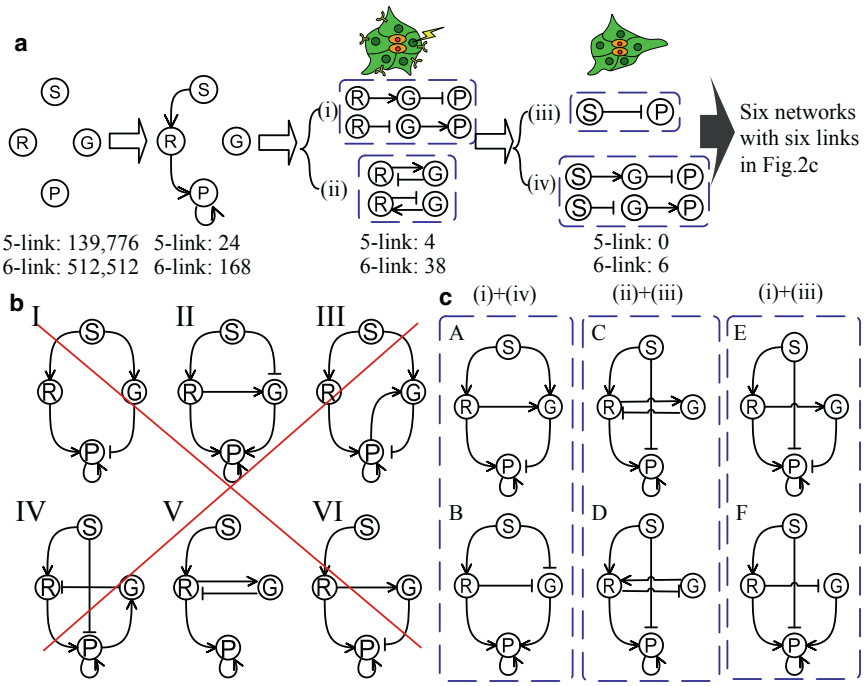

FIGURE 2 Network selection procedure. (*a*) A schematic flow of the qualitative selection procedure is shown. In the first step, the number of candidates is reduced by fixing three links based on other experimental results and by the requirement of at least one input to G and one output from G. In the second and third steps, photoinhibition of Rac and Receptor^{DN} experiments are used, respectively, to further reduce the number of candidates. (*b*) Examples of networks that are excluded using our qualitative selection procedure are shown. Networks I–IV are excluded based on photoinhibition of Rac, whereas the exclusion of V and VI is based on Receptor dominant-negative mutant experiments. Positive interactions are shown as arrows, and negative interactions are shown as bars. For further details, see the text. (*c*) Final results are shown of the qualitative selection procedure, consisting of six networks each with six links. To see this figure in color, go online.

networks, we will not consider topologies with more than six links.

Photoinhibition experiment requires one of four possible interaction motifs

Photoinhibition of Rac in the front cell of a wild-type cluster decreases the probability of that cell protruding but increases the protrusion probability of the other cells in the cluster (13). This information creates two requirements for the network. First, because inhibiting Rac in one cell affects a separate cell's protrusions, there should be an interaction from one cell's R to the other cell's P through the global factor G. This gives us four possible motifs within the network ($R\rightarrow G\rightarrow P$, $R\rightarrow G\rightarrow R\rightarrow P$, $R\rightarrow P\rightarrow G\rightarrow P$, and $R\rightarrow P\rightarrow G\rightarrow R\rightarrow P$), and topologies lacking these motifs can be excluded. An example of such an excluded network is shown in Fig. 2 *b*, I. Second, because inhibition corresponds to a reduction of R, the interaction between R and P through G should be negative. Thus, the product of the signs of the links in these four possible motifs should be negative. We will not consider the possibility of having more than one motif, as this would be redundant, and we are trying to find the minimal number of possible networks. For instance, in the $R\rightarrow G\rightarrow P$ motif, if both links are positive, the net effect from one cell's R to another cell's P will be positive, which contradicts the experiments. Network II in Fig. 2 *b* is an example of a network that can be excluded using this logic. The other three motifs contain more than two links, including a direct link between R and P. Because this $R\rightarrow P$ interaction is required by other experimental results to be positive (31,36), there are also only two possibilities of sign assignments. For example, for the pathway $R\rightarrow G\rightarrow R\rightarrow P$, the sign assignment can only be either $R\overset{+}{\rightarrow}G\overset{-}{\rightarrow}R\overset{+}{\rightarrow}P$ or $R\overset{-}{\rightarrow}G\overset{+}{\rightarrow}R\overset{+}{\rightarrow}P$.

After further analysis, we can rule out both the (a) $R\rightarrow P\rightarrow G\rightarrow P$ and the (b) $R\rightarrow P\rightarrow G\rightarrow R\rightarrow P$ motifs entirely. Two examples including these two excluded motifs are shown in Fig. 2 *b* as III and IV. In these motifs, communication from one cell to another occurs through G using protrusion as an intermediate step. The photoinhibition experiments show that Rac inhibition in one cell leads to a significant increase of protrusions in other cells (see Table 1). Within these motifs, this implies that a protrusion decrease in one cell (P_i) promotes protrusions in other cells (P_j). As the signal S does not change in the photoinhibition experiments, the following logic holds regardless of how S is connected to the other components of the network. So, in the following analysis, when we rule out a motif, we mean that all the networks containing this motif (without redundant motifs) are ruled out. And as $R\rightarrow P$ is fixed to be a positive effect by experiments, we only need to consider the signs of the following two motifs: (a′)$P\rightarrow G\rightarrow P$ and (b′)$P\rightarrow G\rightarrow R$.

As is mentioned in the first paragraph of this section, the phenomenon that inhibiting R_i leads to a decrease in P_j has ruled out the possibility that the effect of motif (a′) is positive. On the other hand, if the effect of motif (a′) is negative (as in Fig. 2 *b*, III), we take $P\overset{+}{\rightarrow}G\overset{-}{\rightarrow}P$ as an example because the logic is similar if we switch the signs. After photoinhibition, the experiments show that more cells display protrusions. This results in an increase in the global factor G because the sign of the link between P and G is positive.

However, because G to P is negative, this will lead to lower P_j, which contradicts the experimental results. So for motif (a′), neither positive nor negative effects match the experiments, and thus, motif (a) is excluded.

We can follow a similar line of logic to exclude motif (b). For this motif, the positive effect is also ruled out in the first paragraph of this section by the phenomenon that inhibiting R_i promotes P_j. Consider the case in which the effect of motif (b′) is negative; we take $P \overset{+}{\rightarrow} G \overset{-}{\rightarrow} R$ (as in Fig. 2 *b*, IV) as an example because switching the signs does not influence the logic. As the number of protrusions increases after photoinhibition, G increases, and thus, R_j decreases. The decrease in R_j, however, will lead to a decrease in P_j decrease, which contradicts the experiments. So motif (b) is also excluded.

Therefore, only two of the four possible motifs remain. These motifs, when taking into account the different sign assignments, result in four possibilities listed in (*i*) and (*ii*) in Fig. 2 *a*. If combined with the fixed links mentioned earlier, they generate four five-link and 38 six-link networks.

Receptor^{DN} experiment requires one of three possible interaction motifs

Wang et al. found that clusters expressing Receptor^{DN} have more protrusions on average than wild-type clusters (13). We model Receptor^{DN} by taking the signal to be zero, meaning that all cells will see identical environments. The comparison to wild-type protrusion numbers will be simpler if we similarly assume that wild-type clusters see a constant but nonzero signal. This will be reasonable in predicting the number of protrusions if the gradient is relatively small. This assumption is supported by measurements of the chemoattractant PVF1 along the border cell cluster's migration path that do not show a steep gradient (12). Furthermore, as we will show below, numerical simulations that take the gradient into account give consistent results.

In the absence of a gradient, the problem is reduced into a homogeneous one, and the average number of protrusions is simply given by $\langle NP \rangle = (\text{number of cells}) \times \text{Prob}$, where Prob is the probability of a protrusion in any cell. The result of the Receptor^{DN} experiment (protrusion number increases when [S] decreases) can be translated into an inequality: $(\mathrm{dProb}/\mathrm{d[S]}) < 0$. Using the total derivative's chain rule, we can express $(\mathrm{dProb}/\mathrm{d[S]})$ in terms of partial derivatives and total derivatives that are experimentally known (see Supporting Materials and Methods for a detailed derivation):

$$\left(1-\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial \mathrm{Prob}}\right)\frac{\mathrm{dProb}}{\mathrm{d[S]}}=\frac{\partial \mathrm{Prob}}{\partial[\mathrm{S}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{S}]}+\left(1-\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial \mathrm{Prob}}\right)\times\frac{\mathrm{dProb}}{\mathrm{d[R]}}\frac{\mathrm{d[R]}}{\mathrm{d[S]}}. \tag{1}$$

Experiments indicate that both total derivatives appearing in the last term are positive: fluorescence resonance energy transfer measurements show that Rac activity increases after adding epidermal growth factor (13), implying $(\mathrm{d[R]}/\mathrm{d[S]}) > 0$, whereas experiments using photoactivatable Rac show that Rac activation induces cell protrusion (24), implying $(\mathrm{dProb}/\mathrm{d[R]}) > 0$. To get $(\mathrm{dProb}/\mathrm{d[S]}) < 0$ in Eq. 1, there are two options:

$$\text{a)}\left(1-\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial \mathrm{Prob}}\right)<0 \text{ and } \left(\frac{\partial \mathrm{Prob}}{\partial[\mathrm{S}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{S}]}\right)>0 \text{ or}$$

$$\text{b)}\left(1-\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial \mathrm{Prob}}\right)>0 \text{ and}\left(\frac{\partial \mathrm{Prob}}{\partial[\mathrm{S}]}+\frac{\partial \mathrm{Prob}}{\partial[\mathrm{G}]}\frac{\partial[\mathrm{G}]}{\partial[\mathrm{S}]}\right)<0.$$

Option a) requires at least three links, namely i) $G \rightarrow P$, $P \rightarrow G$, and $S \rightarrow P$ or ii) $G \rightarrow P$, $P \rightarrow G$, and $S \rightarrow G$ as $(\partial \mathrm{Prob}/\partial[\mathrm{G}])(\partial[\mathrm{G}]/\partial \mathrm{Prob})$, should be nonzero. Combining the links in these two possibilities with the five links already determined above and taking into account possible redundant links lead to networks that have either seven or eight links. Thus, these networks will not be further considered. Option b) requires $(\partial \mathrm{Prob}/\partial[\mathrm{S}]) < 0$ or $(\partial \mathrm{Prob}/\partial[\mathrm{G}])(\partial[\mathrm{G}]/\partial[\mathrm{S}]) < 0$, which can be accomplished with either one or two additional links shown in Fig. 2 *a* (*iii*) and (*iv*). Two examples of excluded networks based on the Receptor^{DN} experiments are shown in Fig. 2 *b* as V and VI.

Final construction of possible topologies

We can now assemble the simplest possible networks by choosing one possible motif mandated by the photoinhibition experiment and one possible motif mandated by the Receptor^{DN} experiment. Specifically, combining motifs from group (*i*) and motifs from group (*iv*) of Fig. 2 *a* results in the two networks A and B shown in the first group of Fig. 2 *c*. Similarly, combining motifs from (*ii*) and (*iii*) leads to networks C and D in Fig. 2 *c*, and combining motifs from groups (*i*) and (*iii*) produces networks E and F. All of these candidate networks have six links: our qualitative selection procedure has ruled out any five-link networks. Choosing motifs from group (*ii*) and group (*iv*) will result in networks with seven links, which we will ignore, as we have found that six-link networks are sufficient to meet our requirements. In the Discussion, we will discuss briefly networks with more links. Finally, it's worth emphasizing again that the above analysis did not specify functional forms of the interactions in the networks; the results will hold regardless of the quantitative details of the interactions.

Simulation results are consistent with selection analysis

In the above analysis of the Receptor^{DN} experiment, we made one key approximation: we assumed that the gradient is shallow enough that protrusion numbers are independent of the gradient. To verify that this approximation does not influence our results and to study the selected networks in a quantitative way, we cast our candidate networks into ODEs and SDEs. Specifically, we have one deterministic equation for R in each cell and a single global deterministic equation for G. These equations are listed in Methods. The equation for the protrusion in cell i, $[P_i]$, is cast as an SDE of the form

$$\frac{d[P_i]}{dt} = F([G], [P_i], [R_i], [S_i]) + \eta(t),$$

where F represents a nonlinear function, and η is a Gaussian noise term with zero mean and SD σ: $\langle \eta(t_1)\eta(t_2)\rangle = 2\sigma^2\delta(t_1 - t_2)$. We have chosen F to be such that P is bistable and can be either in a low state, $[P]_a$, corresponding to a nonprotruding state, or in a high state, $[P]_c$, representing the protruding state (details in Methods). Using a bistable equation for P results in a clear distinction between cells that are protruding (in the high state) and those that are not (low state). To solve the resulting set of 13 coupled equations, we assume that all seven deterministic equations reach equilibrium quickly compared to the timescale of protruding (20–30 min) (13,25). The resulting stable states for R and G can then be substituted into the six SDEs for P_i.

To solve the SDEs, one could in principle use an SDE solver to determine the steady state of the cluster and thus the average protrusion number. However, this would require capturing a large number of transitions between the high and low states. These transitions could become prohibitively slow, especially if the barrier in the bistable protrusion state is large compared to the strength of fluctuations, requiring an excessive amount of computational resources. Therefore, rather than directly solving the SDEs, we treat the problem as a transition between the low state $[P]_a$ and high state $[P]_c$ and compute transition rates between these two states. This is possible because the equation for P can be written as follows:

$$\frac{d[P]}{dt} = -U'([P]) + \eta(t),$$

which is precisely the equation of an overdamped particle in a double-well potential $U([P])$. Thus, we can calculate the transition rates between states $[P]_a$ and $[P]_c$ using the Kramers approximation as follows:

$$k_{on} = \frac{\sqrt{-U''([P]_a)U''([P]_b)}}{2\pi} e^{-\frac{U([P]_b)-U([P]_a)}{\sigma^2}}$$

$$k_{off} = \frac{\sqrt{-U''([P]_c)U''([P]_b)}}{2\pi} e^{-\frac{U([P]_b)-U([P]_c)}{\sigma^2}}.$$

Here, $[P]_b$ is the maximum of the potential $U([P])$ between $[P]_a$ and $[P]_c$ (see also Fig. S1). Note that the Kramers approximation needs the prerequisites that $\Delta U \gg \sigma^2$, where $\Delta U \equiv U([P]_b) - U([P]_{a \text{ or } c})$. When this prerequisite does not hold, we use a more accurate method based on the mean first passage time in a one-dimensional potential to calculate the transition rates (details are in Supporting Materials and Methods). Because the transition rates depend on the ratio of ΔU and σ^2 rather than the specific value of σ^2 and because the value of ΔU is controlled by other parameters, we can fix $\sigma = 0.1$ in our simulation without losing generality.

Given the transition rates between the two states, we can compute the average number of protrusions under many circumstances. Each cell can be either protruding or not; there are then $2^6 = 64$ possible states for a cluster with six cells, and each state α (with $\alpha = 1,2,\ldots,64$) then has a total number of protrusions, NP_α. Associated with each state is a probability W_α, which can be computed using a master equation

$$\frac{dW_\alpha}{dt} = \sum_{\beta=1}^{64} q_{\alpha\beta} W_\beta.$$

Here, q is a matrix with $q_{\alpha\beta}$ ($\alpha \neq \beta$) describing the transition rate from state β to state α, whereas $q_{\alpha\alpha}$ represents the transition rate out of state α, which can be defined as $q_{\alpha\alpha} = -\sum_{\beta=1,\beta\neq\alpha}^{64} q_{\beta\alpha}$. The transition rates are given by the escape rates k_{on} and k_{off} shown above. We will only consider transitions in which one cell changes its state. In the long-time limit, the probabilities tend to the stationary solution, that is $(dW_\alpha/dt) = 0$, resulting in a set of 64 coupled linear equations that can be easily solved (37). Once these probabilities are found, it is straightforward to find the average number of protrusion, computed as $\langle NP\rangle = \sum_\alpha NP_\alpha W_\alpha$, the number of protrusions in the direction of the gradient, or other relevant experimental measurements.

To reduce the computational expense, we study all 24 networks with five links resulting from our initial qualitative selection (shown in the second step of Fig. 2 *a*) but limit ourselves to a subset of 12 six-link networks out of 168 possible networks, including the ones shown in Fig. 2 *c*. In our simulations, we use Latin hypercube sampling to generate 500,000 sets of parameters; because parameters correspond to positive rates, we sample uniformly over the log of the parameters. The parameters and their sampling ranges are listed in Table 2. Then, we apply the same perturbations corresponding to the Receptor^{DN} and photoinhibition experiments and compare the pre- and postperturbation protrusion numbers.

In Fig. 3, we show the sampling results for the five networks illustrated in the upper row. The first four (two

TABLE 2 Range of Model Parameters Used in Sampling Simulations

Name	Range in log10 Space	Comment	Name	Range in log10 Space	Comment
S_{grad}	(−3, 0)		S_{mean}	(−2, 1)	
$basal_{AR}$	(−2, 1)		k_1	(−2, 2)	parameters corresponding to the links, except the three fixed ones;
P^{tot}	(−2, 2)		k_2	(−2, 2)	k_3 is only used in the six-link network
k_{RP}	(−2, 2)		k_3	(−2, 2)	
PI	(−2, 2)	photoinhibition strength	PA		photoactivation strength (only used in fitting, not sampling tests)

five-link and two six-link topologies) were ruled out by our selection procedure, whereas the rightmost network is not excluded by the qualitative selection. The results of the numerical sampling are shown in the middle and lower rows, where each dot corresponds to one set of parameters. The *x* axis represents the number of protrusions in wild-type clusters, whereas the *y* axis represents the protrusion number for the lead cell photoinhibition (*middle row*) and the $Receptor^{DN}$ experiments (*lower row*). The red box shows the experimentally observed protrusion number with the center as the mean and the length and width as the observed SDs. The existence of dots within the red box means that the network is consistent with the experimental data. We see that the first four networks fail to meet the requirements for one of the two experiments, whereas the last network shows parameter combinations that can satisfy both.

We emphasize that in Fig. 3 and in the networks from which we have sampled explicitly, we find results that are consistent with our qualitative models. Models that fail the $Receptor^{DN}$ test qualitatively fail it in sampling, whereas models that can be ruled out by the qualitative selection procedure based on the photoinhibition experiments are also inconsistent with these experiments when analyzing the sampling results. It is worth noting that the points within the red box in the middle row may not be the same sets of parameters as those in the lower row; having points in both red boxes is necessary but not sufficient to be consistent with the experiment. Below, we will fit the parameters for the six selected networks to make sure that we can find at least one set of parameters for each network that is consistent with the experimental results listed in Table 1. More sampling results can be found in Figs. S2–S4, and the sampling results for networks B–F of Fig. 2 *c* are shown in Fig. S4.

Further discrimination between the six minimal networks with six interactions

After finding the six minimal networks with six interactions, the remaining question is how to discriminate between

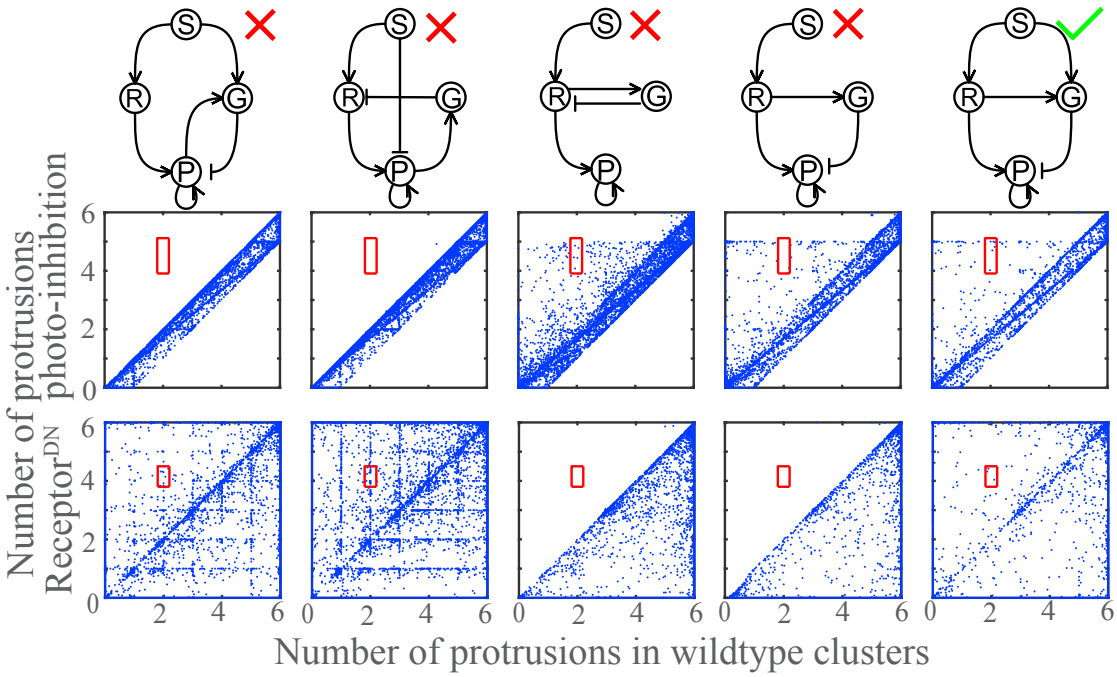

FIGURE 3 Parameter sampling results of five networks. The *x* axes show the number of protrusions for wild-type clusters, and the *y* axes show the number of protrusions for the photoinhibition case (*middle row*) and the $Receptor^{DN}$ case (*lower row*). The rectangular boxes show the region quantitatively consistent with the experimental data, whereas a dot represents the output of one set of parameters. To see this figure in color, go online.

these six networks or these three groups. In addition to the protrusion statistics we have already used, Wang et al. also defined and measured a directionality index defined as $D_{exp} = (\sum_i \vec{P}_i \cdot \vec{d} / \sum_i \|\vec{P}_i\|)$, where $\vec{d}$ is the unit vector in the signal gradient direction, and $\vec{P}_i$ is the protrusion vector for cell i determined from experimental images. To compare this to our model results, we take the direction of the protrusion in the model to be normal to the edge of the round cluster and of length 1 for the protruding state and 0 for the nonprotruding state. Thus, we can define a directionality index for the model using the equation $D_{mod} = (\sum_{i=1}^{6} P_i \cos\theta_i / \sum_{i=1}^{6} P_i)$, which can range from +1 (only the front cell shows a protrusion) to −1 (only the back cell protrudes).

Among the six possible networks in Fig. 2 *c*, networks C, D, E, and F have a negative S→P interaction, whereas networks A and B do not. This negative S→P interaction is determined by the Receptor^{DN} experiments requiring $\langle NP_{Rec^{DN}} \rangle - \langle NP_{wild} \rangle > 0$. However, if the chemoattractant suppresses protrusions, it would tend to make the cluster move against the signal gradient, contradicting the experiments. This is not a critical problem, as there is also a positive effect S→R→P to balance the negative effect. But for networks C, D, E, and F, positive chemotaxis of the cluster is not a robust feature but one requiring fine-tuning of the parameters. In contrast, for networks A and B, the negative effect from S to P is through the global factor G, which has the same value for all the cells, leading to robust chemotaxis. This difference is shown in Fig. 4 *a* by plotting the results of 500,000 samples with the *y* axis as the directionality for the wild-type cluster and the *x* axis as the difference between the average protruding number for Receptor^{DN} clusters and that for wild-type clusters ($\langle NP_{Rec^{DN}} \rangle - \langle NP_{wild} \rangle$). Within Fig. 4 *a*, parameters that pass the Receptor^{DN} standard are in the right half of the plot, whereas parameters with positive directionality are in the top half of the plot; ideally, models should have points in the top right quadrant. Fig. 4 *a* shows more dots in the upper half for networks A, B, and F, which means that it is more likely to obtain good parameters for these three networks compared to the others. More importantly, there are no dots below the *x* axis in networks A and B, meaning that regardless of parameters, these two networks always make the wild-type cluster follow the gradient. In other words, networks A and B robustly have positive directionality. The absence of robust directionality in networks C–F suggests that, if these are correct, further experimental interventions could convert chemotaxis to chemorepulsion.

In addition, we are able to suggest experiments to help further discriminate between the six selected networks. One significant difference between networks C and D of Fig. 2 *c* and the other networks is how the protrusion of one cell is affected by the activation of R in another cell. For networks C and D, this control is through R, which means changing R in one cell influences not only the protrusions of other cells but also the level of R. Therefore, measurement of Rac activation simultaneously with photoactivation would show whether R's activity can be controlled by light treatment on other cells. The results of these experiments can then be used to further discriminate between topologies.

Parameter fitting

To make sure that parameters exist to meet the requirement of all the experiments simultaneously and as a prerequisite to making predictions, we fit 10 parameters for each of the six networks using a simulated annealing algorithm. We fit not only the Receptor^{DN} and photoinhibition experiments but also the other protrusion number data found in Table 1. We also mandate that our fits display signs of directionality that are consistent with the experimental results listed in Table 1. Details of the procedure can be found in the Supporting Materials and Methods. The comparison of protrusion number and directionality between the simulation results with fitted parameters and the experimental data for the six networks under all five conditions listed in Table 1 are shown in Fig. 4 *b*. Clearly, our best-fit parameters are able to produce results that are within similar levels of accuracy for each of the six networks. The values of the fitted parameters for the six networks are listed in Supporting Materials and Methods. It is worth mentioning that we can find more than one set of parameters for each network that can fit the experimental data similarly well, as the parameter space has lots of local minima. Our fitting procedure typically produces multiple parameter sets, and here, for illustrative purposes, we choose the set that has the lowest total error, as defined by our error function. Note, however, that this parameter set might not represent the global minimum and might not be the most optimal one. As a simple robustness check, we perturb one parameter by $\pm 10\%$ while keeping the other parameters constant and quantify the outcomes of the model with the modified parameter. We find that the sensitivity of the model depends on the parameter and on the network topology. For example, for all the six topologies, we find that the results are not sensitive to the parameters S_{mean}, S_{grad}, PI, and PA; a change in these parameters results in a change in the number of protrusions of less than 10%. Changes in other parameters, including k_{RP} and k_{-IP}, lead to larger changes in protrusion numbers for some specific network topologies. Note, however, that our goal is not to find specific parameter values or parameter values that are robust to changes. Instead, we aim to find possible topologies that are consistent with experimental data using strategies that are independent of parameter values. Robustness may be used later as a useful criterion for further distinguishing models (38).

One interesting feature of this figure is that for most conditions, the SDs of the simulations are much larger than those of the experiment. For all the six minimal

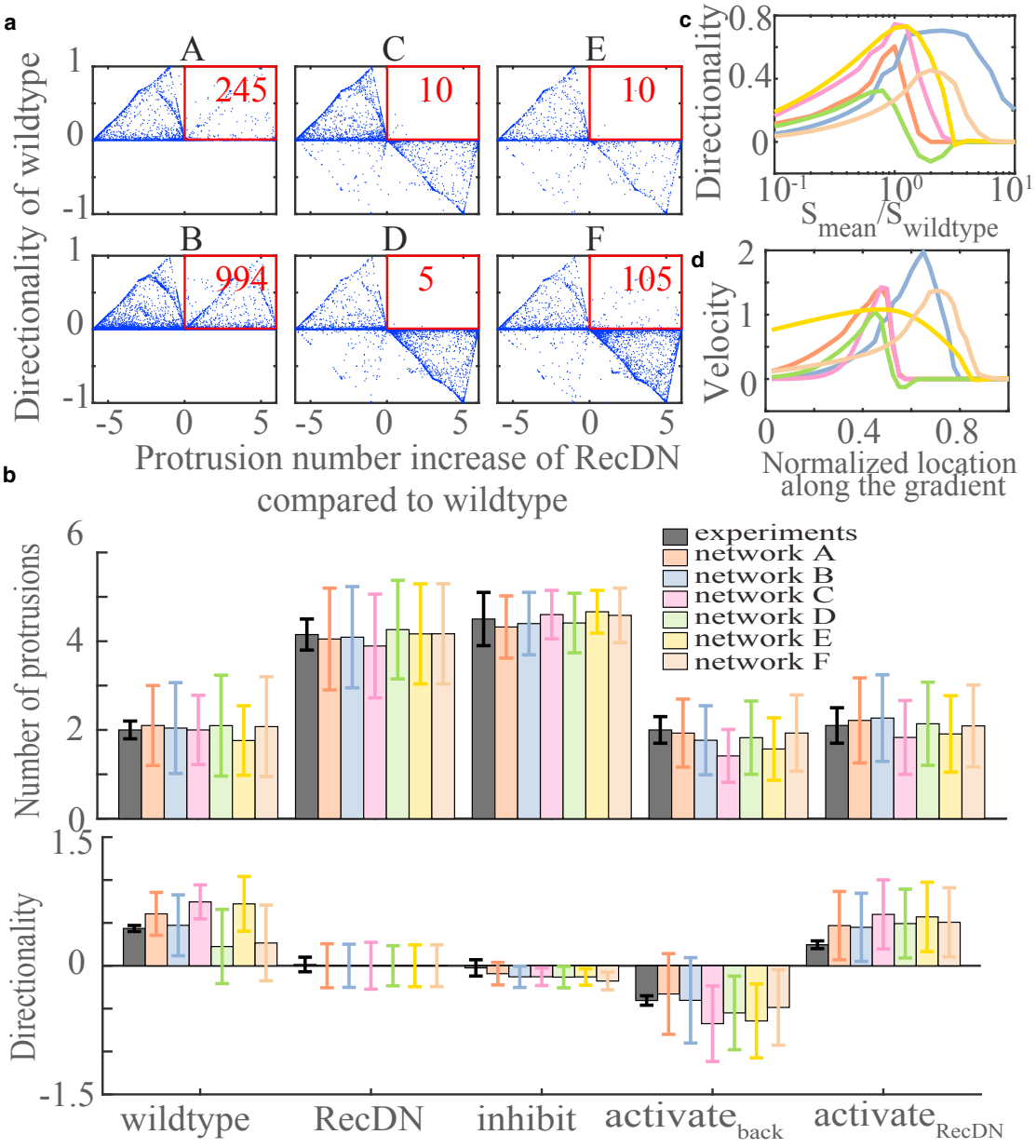

FIGURE 4 Simulation results. (*a*) Parameter sampling results for the six selected networks show am average increase of protrusion number under the $Receptor^{DN}$ condition compared to the wild-type cluster in the *x* axes and the wild-type cluster's directionality in the *y* axes. The square box shows the region that is qualitatively consistent with the experimental results, and the number represents the number of parameter sets (out of 500,000) that produce results that are consistent with the experiments. (*b*) A comparison of the experimental data reported in Table 1 and simulation results for the six selected networks with fitted parameters are shown. A–F correspond to networks A through F in Fig. 2 *c*. Error bars show the SDs. (*c*) Trends of the cluster's directionality are shown for the six selected networks with fitted parameters when the mean signal is changed. (*d*) Trends of the cluster's velocity are shown for the six selected networks with fitted parameters along the migration path. To see this figure in color, go online.

networks, there is no feedback from P to G so that the protrusion of one cell is independent of the protruding states of other cells. With this independence prerequisite, as the signal for every cell is the same in the $Receptor^{DN}$ case, the protruding probabilities for all the cells in a cluster of n cells are given by $\text{Prob} = (4.15/n)$, as the experimental measurement shows an average of 4.15 protrusions under this condition. Now, we can calculate the SD for protrusion number easily, as the probability of having n protrusions is simply $\binom{n}{i}(\text{Prob})^i(1-\text{Prob})^{n-i}$, where $\binom{n}{i}$ is the binomial coefficient. Therefore, $\text{SD} = \sqrt{\sum_{i=0}^{n}(i-4.15)^2\binom{n}{i}(\text{Prob})^i(1-\text{Prob})^{n-i}}$, giving $\text{SD} = 1.13$ for $n = 6$. This is consistent with our simulation result

and is much larger than the experimental SD, which is ~0.35. This means that within a model with six cells in which the probability of a cell protruding is independent of whether other cells have protruded, the SD of the protrusion number cannot be reduced by merely changing the fit parameters. To get an SD close to the experimental result, we have to change either the cell number or the network topology to remove independence. When $n = 5$, $SD \approx 0.84$, and when $n = 4$, to get an average protrusion number close to 4, $Prob = 1$ and thus $SD = 0$. Usually, there are 4–8 border cells in the cluster (27), so one possible reason for the small SD in the experiment is that a significant fraction of cluster samples has only four border cells. The other possibility is to change the independence prerequisite by adding a link from P to G to form negative feedback between P and G on the six selected networks. Networks with a P→G link show reduced variance because the feedback can constrain the average number of protrusions to be very close to a whole number. This is clear in the sampling results of $Receptor^{DN}$ for the first two networks in Fig. 3 showing a pattern with grids at the whole numbers. If the variance is to be tightly constrained in this way, however, we need at least seven interactions for the minimal networks. We give an example of this kind of seven-link network by adding a positive arrow from P to G to network A, shown as network 2 in Fig. S6. The simulation results show that adding this additional feedback can reduce the SD for the number of protrusions but not necessarily for the directionality.

Model predictions

Although we do not know which of the six selected networks is the real one, we can still make some reasonable predictions based on their common simulation results with the fitted parameters. By changing the mean value of the signal while keeping the relative gradient constant, we can get the trends of the directionality. This is shown in Fig. 4 *c*, which reveals that all networks display a maximal directionality near the original mean value of signal. This result is qualitatively robust regardless of network topologies or parameters as long as they meet the requirements of our selecting process. Fig. 4 *c* also shows that the directionality of network D (*green curve*) can become negative for some range of signal strengths. This is consistent with our previous conclusion that the networks, except networks A and B, do not guarantee chemotaxis. We note that a direct quantitative comparison between experimental and model results is difficult. First, experiments have not established the concentration and gradient strength of chemoattractant in vivo. Secondly, we did not consider the length of the protrusion in our model—our protrusions were either on or off. For example, in our model, if all the cells in the cluster have protrusions, the directionality is exactly zero, but in reality, it is possible that the protrusions of the cells at the front half of the cluster are larger, resulting in a positive directionality.

Despite the lack of quantitative comparison, we can still compare our model prediction of Fig. 4 *c* with experimental data. Increasing the mean of signal in a wild-type cluster while keeping the relative gradient constant is similar to PVF1 overexpression. This overexpression was found to decrease the directionality of the cluster (39). Furthermore, experiments have measured the speed profile of the cluster along its path, during which the mean signal is assumed to increase (19). To compare with this experimental data, we define a velocity $V = \sum_i \vec{P}_i \cdot \vec{d}$, where $\vec{d}$ is the unit vector in the signal gradient direction, and $\vec{P}_i$ is the protrusion vector for cell i. Note that in this definition, we do not consider the length of the protrusion nor do we include any mechanical aspect of migration. We assume that the signal profile across the egg chamber is exponential, as suggested in (19), and the signal profile for a cluster at location Lx is $S(x) = S_{mean} e^{-S_{grad}(C-Lx)}(1 + S_{grad} \cos\theta_i)$. Here, L is the total length of the migration path, and x is the normalized location. We set the radius of the cluster to be 1, and by rough measurement of the length of the cluster's travel path from the images in (26), we will take $L = 20$. For C, the location at which the data in (13) are measured, we will take $C = 8$, corresponding to measurements around 1/3–1/2 region of the cluster's travel path (X. Wang, personal communication). S_{mean} and S_{grad} are the parameters fitted in the previous part for the six selected networks and correspond to the mean and relative gradient of the signal at the measured position in the experiments. We plot V vs. x curves for all the six networks in Fig. 4 *d*. We can see that the main trend that velocity first accelerates and then decelerates is consistent with the experimental curves in (19). This behavior emerges from our model, which has been fitted to the perturbation experiments but uses no information from (19) other than the proposed exponential signal profile.

We also examine the role of cell-cell communication in the cluster's chemotaxis by making the global factor G purely local. In the simulation, we use the same sets of parameters we have fitted for the six networks as in Fig. 4 *b* and get negative or zero cluster velocities (Fig. S5 *a*), indicating that cell-cell communication increases the chemotactic ability of the cluster. To verify that these results are not sensitive to the precise choice of parameters, we use the same sets of randomly generated parameters as in Figs. 3 and 4 *a* but reject those with negative velocity. We find that for all six networks, blocking cell-cell communication by making G local results in decreased and sometimes negative cluster velocities (Fig. S5 *b*). In fact, for networks C and D, abolishing global communication always results in clusters that have negative velocities. It should be noted that clusters of mammary epithelial cells with decreased cell-cell communication have been shown to lose their directional response to shallow gradients (20). It would be interesting

to examine this prediction in experiments in which gap junctions between border cells are blocked.

DISCUSSION

In this study, we used mathematical analysis combined with sampling simulations and parameter fitting to determine the simplest network topology that is consistent with a set of in vivo experiments on border cell migration. The mathematical analysis relied solely on qualitative experimental observations and is thus independent of specific functional forms of the interactions between the pathway nodes. Key to this analysis is the fact that specific experimental perturbations, either in the form of photoinhibition, photoactivation, or mutations, resulted in clear phenotypic changes. Using these changes, we were able to reduce the amount of possible simple networks from a very large number to only six.

The complexity of our problem makes application of standard approaches infeasible. A number of existing techniques exist to infer network topologies from experimental data on single cells (20–22). Often, it is possible to gain insight into the topology on the basis of mutant analysis, drug responses, or clever experimentation (40,41). Furthermore, for topologies that contain a limited number of components, it is feasible to perform an exhaustive numerical search through parameter space to determine all possible topologies that are consistent with a particular feature observed in experimental data (35). We have not directly applied these approaches to our problem because inferring network topologies for collective migration is more challenging for multiple reasons. First, cells communicate with each other to sense signal gradient, resulting in a spatially extended system. Second, the output of the system is often a stochastic variable and is only known as an average quantity. In addition to these system-specific challenges, the number of network components might be too large to carry out exhaustive numerical searches. Therefore, we have developed a qualitative perturbation analysis method that can greatly reduce the number of candidate topologies. We then carried out numerical simulations only on the remaining networks for further study.

The final result, i.e., the six possible six-link pathways shown in Fig. 2 *c*, was verified using both sampling simulations and by a fitting procedure that produced parameter sets for which the model results are fully consistent with all experimental data. We should point out that obtaining a satisfactory fit was not necessarily expected. Our fitting procedure takes into account directionality data (Table 1) that was not used in our qualitative selective procedure. Because we found parameter sets for all topologies that were consistent with the experimental data, we are not able to further distinguish between the six networks. However, additional sampling simulations show that some topologies may be more robust than others (Fig. 4 *a*). Furthermore, we propose additional experiments, such as combining light treatment with Rac activity measurement, that should be able to further discriminate between possible topologies.

Our study has a number of limitations. First, our analysis provides the simplest four-node topologies that are able to reproduce the experimental data, which turn out to be topologies with six links. Within this class of four-node networks, it is likely that topologies with more than six links can also be consistent with the data. Determining all possible four-node topologies, however, would be computationally prohibitive. Second, it is possible that more complex networks containing more nodes are also able to conform to the data. We show the simulation results for two networks with seven links in Fig. S6. Network 1 is obtained from a combination of the motifs in groups (*ii*) and (*iv*) in Fig. 2 *a* and adds negative feedback from G to R in network A (Fig. 2 *c*). We again fitted the parameters of this new network and found that the simulation results are similar to the ones obtained in our six-link networks (Fig. S6 and Fig. 4 *b*). Network 2 is obtained by adding a positive interaction from P to G in network A (Fig. 2 *c*). Simulations of this network show that the error bars for the number of protrusions are smaller than the ones for the six-link networks (Fig. S6 and Fig. 4 *b*). The results of network 2 show that new phenomena (smaller error bars in the example) are possible with more links. We do not view the minimal networks with four nodes and six links as the final word on the subject but suggest them as a reasonable start for further research, much as minimal LEGI models have proved useful for studying chemotaxis in other contexts. In summary, our goal in this study was to find minimally complex networks that can then be further selected with additional experiments. It is conceivable that these experiments, which to our knowledge are new, will result in the elimination of all six networks, in which case more complex networks or networks with more links need to be considered. Procedures similar to the ones we used here should then be able to provide possible signaling-pathway architectures.

Despite these limitations, our results provide valuable information about the multicellular signal-processing network in the collective migration of border cells, showing the defect of the multicellular LEGI model (network I in Fig. 2 *b*) that was studied in (20–23) and narrowing the topology to only six possibilities. These models may be interesting candidate models to further study collective migration in the border cell system as well as in other systems. For example, in (14,42), Rab11 and Cdc42's functions in cell-cell communication are studied using mutations and light treatment of Rac. With their results, we can check which roles they may play in our topologies, and further discrimination of the six topologies may be possible; otherwise, we may need a larger network including some new components. The six topologies that we have found are also potential candidates to study in contexts in which minimal multicellular LEGI models have been previously

applied (20–23). These topologies may potentially provide more gradient-sensing accuracy than the LEGI scheme (20,21) or may have different chemotactic dependence on cluster sizes (19,22,23). In addition, the photoactivatable Rac has been used in other cell types, such as mesendoderm cells (33) and neural crest cells (34), that also show collective cell migration. Application of our approach combined with Rac activation experiments in these systems could show whether the topologies we have found for the border cell cluster migration are robust across different cell types or whether different topologies are optimal for differential biological circumstances. We should point out that if photoactivatable Rac, an excellent tool for the reversible manipulation of proteins, is not practicable, traditional genetic methods may also provide data that can be analyzed by our qualitative approach. For example, in (43), DNA-programmed epithelial cell assembly is used to study multicellular protrusions when Ras is activated in all or some of the cells. The comparison of homogeneous and heterogeneous Ras activity can provide a similar kind of information as that of the Receptor^{DN} (homogeneous inhibition) and the local light treatment of Rac (heterogeneous activation or inhibition) experiments we use in our work. Thus, our work provides a starting point for future research on signaling networks of collective chemotaxis, and the methods we have developed here may be applied in other collective migrating systems as well.

SUPPORTING MATERIAL

Supporting Materials and Methods, six figures, and two tables are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(18)30466-1.

AUTHOR CONTRIBUTIONS

H.Y. designed the research, performed the research, analyzed the data, and wrote the manuscript. B.A.C. and W.-J.R. designed the research, analyzed the data, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Denise Montell and Xiaobo Wang for useful conversations.

This work was supported by National Institutes of Health grant number P01 GM078586.

SUPPORTING CITATIONS

Reference (44) appears in the Supporting Material.

REFERENCES

1. Camley, B. A., and W. J. Rappel. 2017. Physical models of collective cell motility: from cell to tissue. *J. Phys. D Appl. Phys.* 50:113002.
2. Hakim, V., and P. Silberzan. 2017. Collective cell migration: a physics perspective. *Rep. Prog. Phys.* 80:076601.
3. Weijer, C. J. 2009. Collective cell migration in development. *J. Cell Sci.* 122:3215–3223.
4. Friedl, P., and D. Gilmour. 2009. Collective cell migration in morphogenesis, regeneration and cancer. *Nat. Rev. Mol. Cell Biol.* 10:445–457.
5. Rørth, P. 2009. Collective cell migration. *Annu. Rev. Cell Dev. Biol.* 25:407–429.
6. Mayor, R., and S. Etienne-Manneville. 2016. The front and rear of collective cell migration. *Nat. Rev. Mol. Cell Biol.* 17:97–109.
7. Camley, B. A., and W. J. Rappel. 2017. Cell-to-cell variation sets a tissue-rheology-dependent bound on collective gradient sensing. *Proc. Natl. Acad. Sci. USA.* 114:E10074–E10082.
8. Malet-Engra, G., W. Yu, …, L. Dupré. 2015. Collective cell motility promotes chemotactic prowess and resistance to chemorepulsion. *Curr. Biol.* 25:242–250.
9. Theveneau, E., L. Marchant, …, R. Mayor. 2010. Collective chemotaxis requires contact-dependent cell polarity. *Dev. Cell.* 19:39–53.
10. Poujade, M., E. Grasland-Mongrain, …, P. Silberzan. 2007. Collective migration of an epithelial monolayer in response to a model wound. *Proc. Natl. Acad. Sci. USA.* 104:15988–15993.
11. Sepúlveda, N., L. Petitjean, …, V. Hakim. 2013. Collective cell motion in an epithelial sheet can be quantitatively described by a stochastic interacting particle model. *PLoS Comput. Biol.* 9:e1002944.
12. Cai, D., S. C. Chen, …, D. J. Montell. 2014. Mechanical feedback through E-cadherin promotes direction sensing during collective cell migration. *Cell.* 157:1146–1159.
13. Wang, X., L. He, …, D. J. Montell. 2010. Light-mediated activation reveals a key role for Rac in collective guidance of cell movement in vivo. *Nat. Cell Biol.* 12:591–597.
14. Ramel, D., X. Wang, …, G. Emery. 2013. Rab11 regulates cell-cell communication during collective cell movements. *Nat. Cell Biol.* 15:317–324.
15. Stonko, D. P., L. Manning, …, B. E. Peercy. 2015. A mathematical model of collective cell migration in a three-dimensional, heterogeneous environment. *PLoS One.* 10:e0122799.
16. Abercrombie, M., and E. J. Ambrose. 1962. The surface properties of cancer cells: a review. *Cancer Res.* 22:525–548.
17. Carmona-Fontaine, C., H. K. Matthews, …, R. Mayor. 2008. Contact inhibition of locomotion in vivo controls neural crest directional migration. *Nature.* 456:957–961.
18. Camley, B. A., J. Zimmermann, …, W. J. Rappel. 2016. Emergent collective chemotaxis without single-cell gradient sensing. *Phys. Rev. Lett.* 116:098101.
19. Cai, D., W. Dai, …, D. J. Montell. 2016. Modeling and analysis of collective cell migration in an in vivo three-dimensional environment. *Proc. Natl. Acad. Sci. USA.* 113:E2134–E2141.
20. Ellison, D., A. Mugler, …, A. Levchenko. 2016. Cell-cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis. *Proc. Natl. Acad. Sci. USA.* 113:E679–E688.
21. Mugler, A., A. Levchenko, and I. Nemenman. 2016. Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proc. Natl. Acad. Sci. USA.* 113:E689–E695.
22. Camley, B. A., J. Zimmermann, …, W. J. Rappel. 2016. Collective signal processing in cluster chemotaxis: roles of adaptation, amplification, and co-attraction in collective guidance. *PLoS Comput. Biol.* 12:e1005008.
23. Varennes, J., B. Han, and A. Mugler. 2016. Collective chemotaxis through noisy multicellular gradient sensing. *Biophys. J.* 111:640–649.
24. Wu, Y. I., D. Frey, …, K. M. Hahn. 2009. A genetically encoded photoactivatable Rac controls the motility of living cells. *Nature.* 461:104–108.
25. Prasad, M., and D. J. Montell. 2007. Cellular and molecular mechanisms of border cell migration analyzed using time-lapse live-cell imaging. *Dev. Cell.* 12:997–1005.

26. Montell, D. J., W. H. Yoon, and M. Starz-Gaiano. 2012. Group choreography: mechanisms orchestrating the collective movement of border cells. *Nat. Rev. Mol. Cell Biol.* 13:631–645.
27. Montell, D. J. 2003. Border-cell migration: the race is on. *Nat. Rev. Mol. Cell Biol.* 4:13–24.
28. Rørth, P. 2002. Initiating and guiding migration: lessons from border cells. *Trends Cell Biol.* 12:325–331.
29. Duchek, P., K. Somogyi, …, P. Rørth. 2001. Guidance of cell migration by the Drosophila PDGF/VEGF receptor. *Cell.* 107:17–26.
30. Duchek, P., and P. Rorth. 2001. Guidance of cell migration by EGF receptor signaling during Drosophila oogenesis. *Science.* 291:131–133.
31. Ridley, A. J., M. A. Schwartz, …, A. R. Horwitz. 2003. Cell migration: integrating signals from front to back. *Science.* 302:1704–1709.
32. Inaki, M., S. Vishnu, …, P. Rørth. 2012. Effective guidance of collective migration based on differences in cell states. *Proc. Natl. Acad. Sci. USA.* 109:2027–2032.
33. Sonavane, P. R., C. Wang, …, D. W. DeSimone. 2017. Mechanical and signaling roles for keratin intermediate filaments in the assembly and morphogenesis of Xenopus mesendoderm tissue at gastrulation. *Development.* 144:4363–4376.
34. Scarpa, E., A. Szabó, …, R. Mayor. 2015. Cadherin switch during EMT in neural crest cells leads to contact inhibition of locomotion via repolarization of forces. *Dev. Cell.* 34:421–434.
35. Ma, W., A. Trusina, …, C. Tang. 2009. Defining network topologies that can achieve biochemical adaptation. *Cell.* 138:760–773.
36. Murphy, A. M., and D. J. Montell. 1996. Cell type-specific roles for Cdc42, Rac, and RhoL in Drosophila oogenesis. *J. Cell Biol.* 133:617–630.
37. Van Kampen, N. G. 1992. Stochastic Processes in Physics and Chemistry. Elsevier, New York.
38. Alon, U. 2007. An Introduction to Systems Biology: Design Principles of Biological Circuits. Chapman and Hall/CRC, Boca Raton, FL.
39. Bianco, A., M. Poukkula, …, P. Rørth. 2007. Two distinct modes of guidance signalling during collective migration of border cells. *Nature.* 448:362–365.
40. Takeda, K., D. Shao, …, R. A. Firtel. 2012. Incoherent feedforward control governs adaptation of activated ras in a eukaryotic chemotaxis pathway. *Sci. Signal.* 5:ra2.
41. Wang, C. J., A. Bergmann, …, A. Levchenko. 2012. Diverse sensitivity thresholds in dynamic signaling responses by social amoebae. *Sci. Signal.* 5:ra17.
42. Colombié, N., V. Choesmel-Cadamuro, …, D. Ramel. 2017. Non-autonomous role of Cdc42 in cell-cell communication during collective migration. *Dev. Biol.* 423:12–18.
43. Liu, J. S., J. T. Farlow, …, Z. J. Gartner. 2012. Programmed cell-to-cell variability in Ras activity triggers emergent behaviors during mammary epithelial morphogenesis. *Cell Rep.* 2:1461–1470.
44. Press, W. H. 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing. Cambridge University Press, Cambridge, UK.