Leader cells in collective chemotaxis: Optimality and trade-offs

Austin Hopkins
Department of Physics & Astronomy, Johns Hopkins University, Baltimore, Maryland 21218, USA

Brian A. Camley*
Department of Physics & Astronomy and Department of Biophysics, Johns Hopkins University, Baltimore, Maryland 21218, USA

(Received 23 May 2019; published 30 September 2019)

Clusters of cells can work together in order to follow a signal gradient, chemotaxing even when single cells do not. Cells in different regions of collectively migrating neural crest streams show different gene expression profiles, suggesting that cells may specialize to leader and follower roles. We use a minimal mathematical model to understand when this specialization is advantageous. In our model, leader cells sense the gradient with an accuracy that depends on the kinetics of ligand-receptor binding, while follower cells follow the cluster's direction with a finite error. Intuitively, specialization into leaders and followers should be optimal when a few cells have more information than the rest of the cluster, such as in the presence of a sharp transition in chemoattractant concentration. We do find this—but also find that high levels of specialization can be optimal in the opposite limit of very shallow gradients. We also predict that the best location for leaders may not be at the front of the cluster. In following leaders, clusters may have to choose between speed and flexibility. Clusters with only a few leaders can take orders of magnitude more time to reorient than all-leader clusters.

DOI: 10.1103/PhysRevE.100.032417

I. INTRODUCTION

Eukaryotic cells commonly chemotax, moving in response to a chemical gradient, to locate wounds and move in a developing embryo. Clusters of cells often chemotax differently than single cells, cooperating to improve their sensing abilities [1–5]. Gene profiling of neural crest cells has shown that cells in different regions of the migrating stream show varying gene expression, including in the genes responsible for cell receptors [6]. Experiments on the zebrafish lateral line primordium demonstrated that adding chemosensing cells to a cluster with reduced chemosensing abilities can help restore migration [7,8]. These results suggest that cells may cooperate by specializing into leaders who sense the chemical gradient, while others follow [2,3,6,9], as previously hypothesized by others [10]. The leaders we describe need not be at the tissue front, differing from the definition of leaders commonly used in wound healing [11]. The term "leader" is also often ambiguously used to describe cells responsible for any of force generation, steering, or sensing [12]; here we focus on sensing.

We expect specialization to leaders and followers to be most important if there is a large difference between the information different cells have about the gradient orientation, as in sharp transitions. This may occur when cell clusters follow a gradient that is "self-generated," i.e., when cells near the rear degrade or sequester chemoattractant [9,13–20], allowing a cell cluster to migrate over distances much longer than its size during development [9,16,19] and cancer metastasis [17,20].

We develop a minimal model of cluster chemotaxis with leaders and followers. Past theoretical and experimental work on self-generated gradients [14,16] has shown that the concentration profile takes a sigmoidal shape, with high concentrations in front of the cluster and lower concentrations at the cluster's rear. Rather than model the details of a specific system's self-generated gradient, we focus on a how a cell cluster behaves in an applied gradient shape and vary the imposed concentration gradient. Specializing improves the chemotactic velocity in both sharp transitions and near-linear gradients of chemoattractant but is not always beneficial. Although it is often assumed that cells near the front of a cluster are best positioned to lead, we find that, depending on the gradient shape, cells near the middle or back of the cluster can have more information about the gradient direction. Specialization not only impacts the migration speed, but also strongly increases the cluster's reorientation time.

II. MODEL AND METHODS

We parametrize the chemoattractant profile:

$$C(x) = \frac{1}{2} C_{\max}\left[1 + \tanh\left(\frac{x}{h}\right)\right]. \tag{1}$$

This function interpolates between steplike gradients and near-linear gradients depending on h, the scale of the transition from 0 to $C_{\max}$ (Fig. 1). We measure lengths in units of the cell diameter, so $h = 1$ is nearly steplike on the scale of a cell cluster. While the cluster moves in the xy plane, the position x in Eq. (1) is measured *relative to the lead cell at* $x = 0$; the cluster does not move relative to the gradient even as it moves in the laboratory frame. This is consistent with measurements of Sdf1 gradients in the zebrafish lateral line, which reach a

*Corresponding author: bcamley@jhu.edu

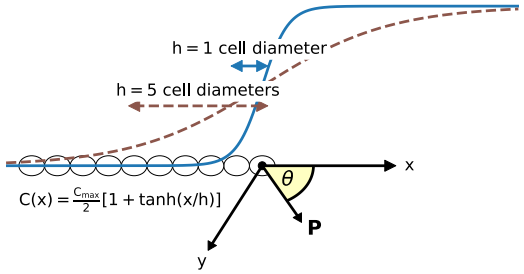

FIG. 1. Geometry. The cell cluster is a rigid train with its front defining the point $x = 0$. The cells move in the xy plane, with polarity $\mathbf{P} = (\cos\theta, \sin\theta)$. h sets the width of the transition regime of the gradient in units of the cell diameter.

steady state in which they maintain their shape and move with the cluster [9,16].

Leader cells make a measurement of the chemoattractant gradient direction. Earlier theory [21–26] and experiments [27–29] have established that a single cell sensing a chemical gradient is often limited in accuracy by the stochasticity of ligand-receptor binding [27]. We extend the model of [22,23], assuming that leaders measuring the chemoattractant orientation have an angular error limited by ligand-receptor interactions. In the shallow gradient limit, this error Δ_ϕ^2 is [22,23]

$$\Delta_\phi^2 \approx \frac{8(C_0 + K_d)^2}{N_r p^2 C_0 K_d} \quad \text{(leaders in shallow gradients)}, \qquad (2)$$

where $C_0 = C(x)$ is the mean concentration near the cell, K_d is the ligand-receptor dissociation constant, N_r is the number of receptors, and $p = \frac{1}{C_0}|\nabla C|$ is the percentage change in concentration across the cell. This shallow-gradient assumption may fail at sharper transitions (e.g., $h = 1$), so we determine the uncertainty Δ_ϕ^2 without this approximation by numerical integration (Appendix A). We plot the leader angular error Δ_l as a function of the position within the cluster in Fig. 2. The follower uncertainty Δ_f is independent of position—i.e., we assume that this noise arises from a process independent of chemosensing.

To relate uncertainties in sensing to cell motion, we describe cells as actively moving and reorienting particles. Each cell i has an orientation θ_i, corresponding to the cell's being polarized in the direction $\mathbf{P}_i = (\cos\theta_i, \sin\theta_i)$. Here, leader cells align with the chemoattractant direction ($\hat{x}$ or $\theta = 0$), while follower cells follow the cluster's direction θ_c [Eq. (5)]:

$$\frac{d\theta_\ell}{dt} = -\frac{1}{\tau_\ell}\theta_\ell + \sigma_\ell \xi_\ell(t) \quad \text{(leaders)}, \qquad (3)$$

$$\frac{d\theta_f}{dt} = -\frac{1}{\tau_f}(\theta_f - \theta_c) + \sigma_f \xi_f(t) \quad \text{(followers)}. \qquad (4)$$

σ and τ for the leaders and followers depend on their accuracies (Appendix B), $\Delta_{l,f}$, and may vary depending on the cell position. $\xi_i(t)$ is Gaussian white noise with $\langle \xi_i(t) \rangle = 0$, $\langle \xi_i(t)\xi_j(t') \rangle = \delta(t - t')\delta_{ij}$. The angles θ_ℓ and $\theta_f - \theta_c$ are interpreted *modulo* 2π; we simulate Eqs. (3) and (4) by the Euler-Maruyama method with $\Delta t = 0.01$ (simulation and numerical details are presented in Appendix C).

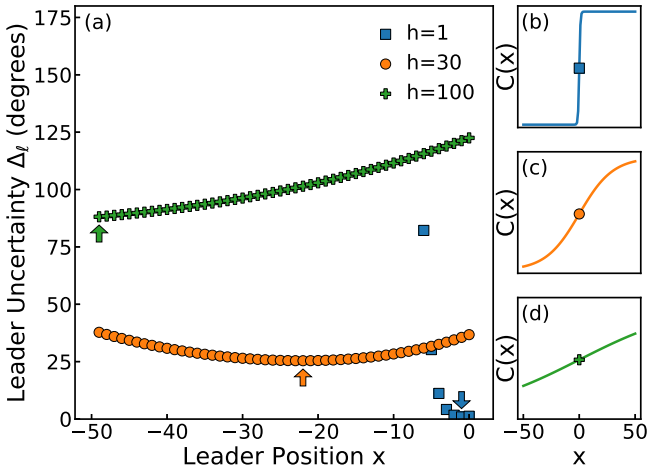

FIG. 2. Directional uncertainty of leaders. (a) Uncertainty Δ_ϕ in a leader's measurement of the gradient direction as a function of leader position x for sharp ($h = 1$), intermediate ($h = 30$), and shallow ($h = 100$) gradients. Arrows indicate the minima of the curves, showing where the first leader will be added in a 50-cell train. $C(x)$ is shown for (b) $h = 1$, (c) $h = 30$, and (d) $h = 100$. $N_r = 70\,000$ and $C_{\max} = 2K_d$ here and throughout the paper.

Collective migration is induced by having follower cells align with the cluster velocity direction θ_c, defined by

$$\mathbf{V}_{\text{cluster}} = |\mathbf{V}_{\text{cluster}}|(\cos\theta_c, \sin\theta_c) = \frac{1}{N}\sum_i \mathbf{P}_i. \qquad (5)$$

The cluster center-of-mass velocity is $\mathbf{V}_{\text{cluster}} = \frac{1}{N}\sum_i \mathbf{P}_i$ when cells are mechanically linked, and cell i would travel at velocity $\mathbf{P}_i$ in the absence of mechanical linkage [30]. The alignment of follower cells with the cluster orientation [Eq. (4)] is a variant of the "self-alignment" [31] mechanism of Szabo *et al.* [32] and others [33–35] who showed that when cells align their polarity with their velocity, mechanical interactions between cells cause cells to align and migrate collectively. Our follower model is precisely that of [32] if the cell velocity and cluster velocity are equal, i.e., the cluster is rigid.

We choose a dependence $\sigma(\Delta)$ and $\tau(\Delta)$ so that Δ^2 controls the variance of the polarity angle and the correlation time $T_{\mathbf{P}}$ of a single cell's polarity $\mathbf{P}_i$ is equal to 1 at all Δ (Appendix B). This corresponds to the assumption that at all levels of uncertainty Δ, a cell reorients at the same time scale, which we choose as our unit time.

III. RESULTS AND DISCUSSION

A. Optimal leadership strategies depend on the chemoattractant profile and follower accuracy

A cluster may, depending on the chemoattractant profile $C(x)$ and the accuracy of its followers Δ_f, improve its mean velocity in the chemoattractant direction $\langle V_x \rangle$ by specializing to leader and follower roles (Fig. 3) [36]. This is similar to results from a more complex model describing leaders and followers and explicitly modeling the process of chemoattractant degradation [37]. We do not assume a specific mechanism that determines which cells lead and which ones follow. Rather, we explore how a cluster behaves as the number of leaders

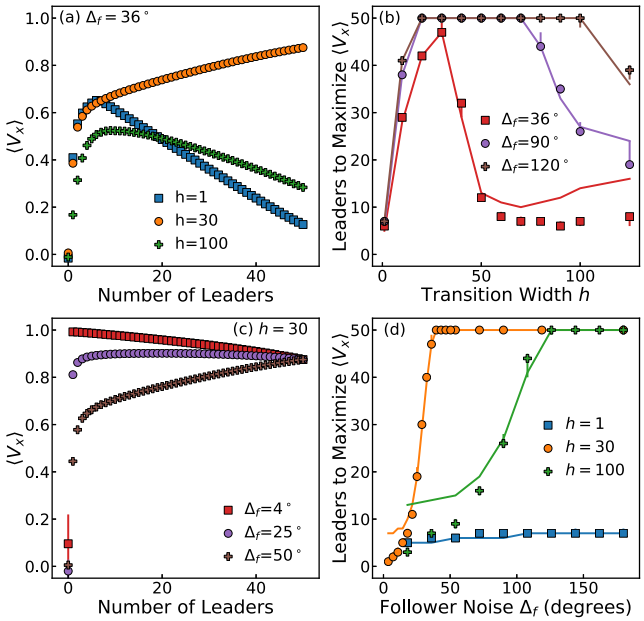

FIG. 3. Cluster velocity depends on leader number. Trains of $N = 50$ cells. Symbols are simulations with 95% confidence intervals from bootstrapping; lines are the independent follower approximation (Appendix D). (a) For sharp ($h \sim 1$) and wide ($h \sim 100$) gradients, the cluster can migrate substantially more quickly in the gradient direction with fewer leaders. (b) The number of leaders that maximizes $\langle V_x \rangle$, $N_\ell^{\max}$, increases, then decreases as a function of the gradient transition width h. (c) At fixed $h = 30$, raising the follower noise Δ_f lowers $\langle V_x \rangle$ at leader fractions of less than 1. (d) $N_\ell^{\max}$ transitions from low to high for medium and wide gradients but remains relatively constant for sharp gradients as the follower noise is increased.

changes. Here, and elsewhere, we add leaders from most to least accurate. We initially study a linear train of cells (Fig. 1), both for simplicity and as the most relevant geometry for narrow, extended systems like the zebrafish lateral line [14].

For sharp transitions ($h = 1$), $\langle V_x \rangle$ first increases and then decreases as we increase the number of leaders, reaching a maximum at $N_\ell^{\max} \approx 6$ [Fig. 3(a)]. Consistent with our intuition that specialization will be most effective when transitions are sharp, $\langle V_x \rangle$ increases monotonically in the number of leaders for the wider transition ($h = 30$). By contrast, in the nearly flat concentration profile of $h = 100$, the chemotactic velocity is maximized by having only a few cells be leaders [Fig. 3(a)].

To better understand how the optimal number of leaders depends on the chemoattractant profile, we study $N_\ell^{\max}$, the number of leaders which maximizes $\langle V_x \rangle$ [Fig. 3(b)]. $N_\ell^{\max}$ is small at sharp gradients and initially increases as the transition size h increases—reflecting that for sharp transitions, leaders even a few cells away from the transition have extremely high levels of uncertainty and will not increase $\langle V_x \rangle$. We would expect that further increasing h places more cells in the transition region and would increase $N_\ell^{\max}$ monotonically. Instead, we see that $N_\ell^{\max}$ decreases at large h as the profile $C(x)$ becomes nearly linear.

This apparently counterintuitive result can be understood directly from the leader uncertainty Δ_ℓ as a function of the cell position (Fig. 2). For gradients with a sharp transition regime ($h = 1$), the leader uncertainty steeply increases for cells farther away from the train front. At wider gradients ($h = 30$), there is a smaller difference between the best and the worst leader. In fact, the most effective leaders tend to be in the middle of the cluster. As the gradient becomes near-linear ($h = 100$), instead of having cells with equal levels of uncertainty, cells near the back of the cluster have a significantly lower uncertainty. This is because in linear gradients, the percentage change across the cell p is maximized farther from the transition, where the baseline concentration is lower, and p limits the accuracy [Eq. (2)]. Specialization is rewarded at large h *because* there is a relevant difference in information available across the cluster.

Specialization relies on followers accurately using information from the leaders; the follower noise Δ_f can qualitatively change how $\langle V_x \rangle$ depends on the number of leaders [Fig. 3(c)]. If follower noise is very low ($\Delta_f = 4^\circ$), one leader can guide the cluster more effectively than when all the cells are leaders. For larger follower noises, $\langle V_x \rangle$ is maximized when every cell is a leader [$\Delta_f = 50^\circ$; Fig. 3(c)]. This leads to an even more dramatic change in $N_\ell^{\max}$: at $h = 30$, there is a rapid switch from $N_\ell^{\max} = 1$ to $N_\ell^{\max} = N$ [Fig. 3(d)] as the follower noise is increased. However, this switching depends on the gradient width h. For the sharp $h = 1$ profile, $N_\ell^{\max}$ does not change much as Δ_f increases; having a small number of leaders is a *robust* strategy in sharp transitions. In the wider $h = 30$ and $h = 100$ gradients, increasing the follower noise level causes the number of leaders which maximizes $\langle V_x \rangle$ to switch from low to high. Again, this can be understood by referring to Fig. 2. When there is a large difference between the best and the worst sensors ($h = 1$), the magnitude of follower noise is relatively unimportant: Δ_f is usually larger than Δ_ℓ for the few well-informed cells but smaller than Δ_ℓ for the bulk of the cells. By contrast, for $h = 30$, most cells have roughly the same amount of information about the gradient direction, and changing Δ_f can rapidly switch between $\Delta_f > \Delta_\ell$ for all cells, in which case it is optimal to have all cells be leaders, and $\Delta_f < \Delta_\ell$ for all cells, when as few cells as possible should lead.

Though our model of Eqs. (3) and (4) has a complex long-range collective interaction, we can quantitatively understand Fig. 3 with a much simpler independent follower model (solid lines in Fig. 3; see Appendix D). Our independent follower model assumes that the follower error $\theta_{\rm rel} \equiv \theta_f - \theta_c$ is independent of θ_c, and also assumes $\theta_c \approx \theta_L$, where θ_L is the angle of only the leaders, $\mathbf{P}_L = \frac{1}{N_\ell} \sum_\ell \mathbf{P}_\ell$. Effectively, each follower then independently follows the leader cells. The independent follower approximation is most effective at high levels of follower noise Δ_f (where follower-follower correlations are shorter-lived and less important) and sharp gradients (low h).

B. Cluster reorientation and leader strategy

Because of the correlations between followers, a collectively sensing cluster can be highly persistent—even if it is moving in an incorrect direction. To understand this persistence, and the time it takes to reorient, we compute

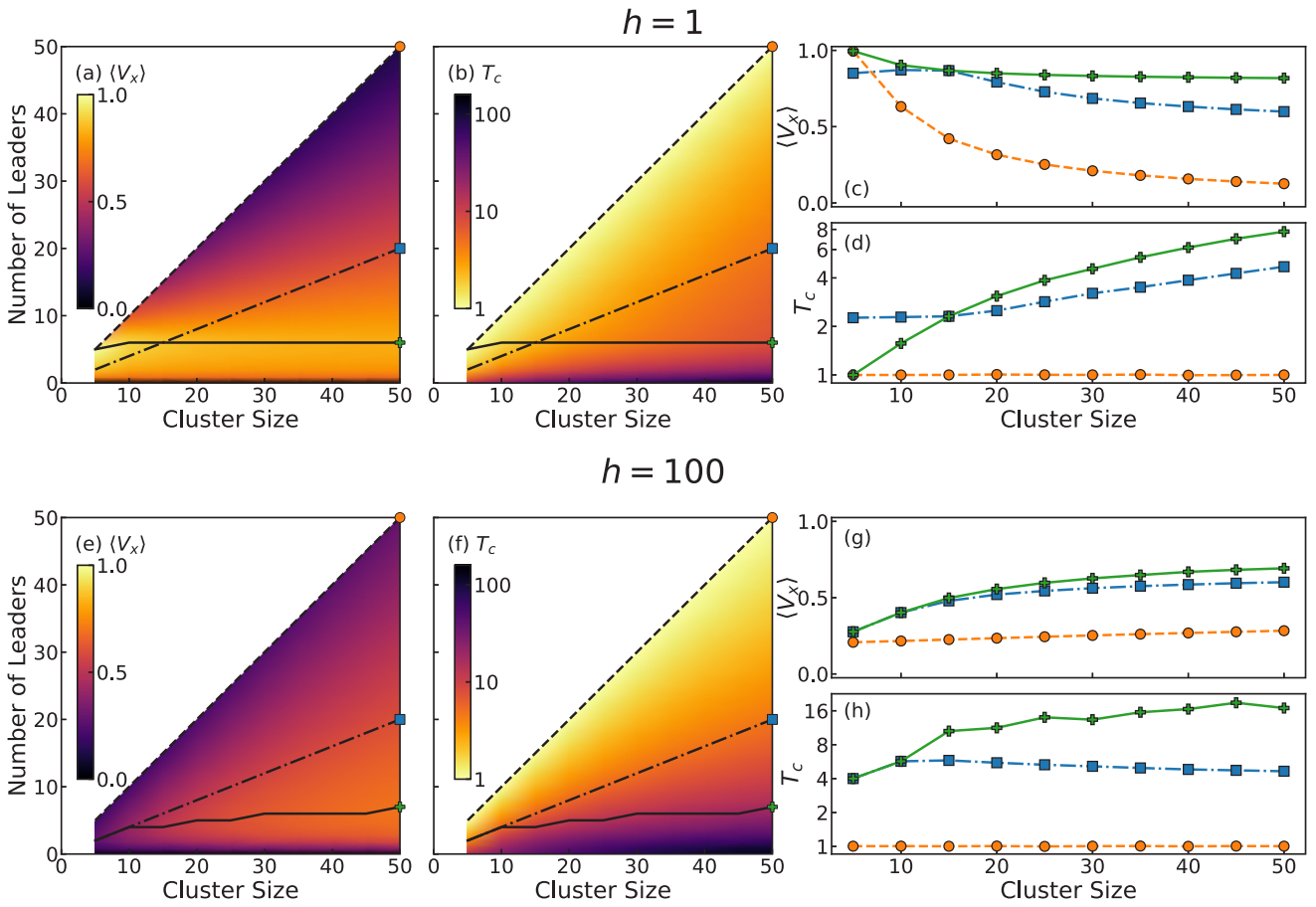

FIG. 4. Cluster size and leader strategy alter the mean velocity and correlation time. Trains of $N = 50$ cells in (a–d) a sharp $h = 1$ gradient and (e–h) a wide $h = 100$ gradient with $\Delta_f = 36°$. Lines represent strategies of allocating leaders (not the independent follower approximation as in Fig. 3). The solid line (green crosses) corresponds to the number of leaders which maximizes $\langle V_x \rangle$. The dashed line (orange circles) corresponds to a leader fraction of 1 and the dot-dashed line (blue squares) corresponds to a leader fraction of 2/5. (a, e) Mean velocity in the gradient direction $\langle V_x \rangle$ as a function of cluster size and number of leaders. The corresponding slices are plotted in (c) and (g). (b, f) Cluster correlation time T_c as a function of cluster size and number of leaders. At a fixed cluster size, this time can vary over two orders of magnitude from low to high leader number. The corresponding slices are plotted in (d) and (h). The 95% confidence intervals from bootstrapping in (c), (d), (g), and (h) are smaller than the symbol sizes.

the velocity autocorrelation function $\langle \delta \mathbf{V}(t) \cdot \delta \mathbf{V}(t') \rangle$ and fit it to an exponential to find the cluster's correlation time T_c (Appendix C). A short correlation time could be advantageous if cell clusters need to rapidly change direction (e.g., metastasizing clusters [38]), while long correlation times could be preferred for cell clusters traveling in consistent directions that must resist perturbations in the concentration profile (e.g., the zebrafish primordium). Experimentally, larger cell clusters exhibit slower reorientation in electric fields [39] and generally slower spontaneous reorientation in confinement [40], so we study both the effect of the cluster size and the number of leaders.

The cluster velocity and correlation time vary significantly with the cluster size N and number of leaders (Fig. 4). For sharp gradients ($h = 1$), specialization $N_\ell < N$ improves $\langle V_x \rangle$ for all $N > 5$; chemotactic velocities decrease sharply for the all-leader ($N_\ell = N$) case at larger cluster sizes, as more and more uninformed cells are leading. Most strikingly, in larger clusters, as the number of leaders is decreased from $N_\ell = N$, the correlation time T_c increases over two orders of magnitude [Fig. 4(b)]. These changes are reminiscent of those observed in a much more complex model of [41]. T_c also increases with smaller leader numbers at large h [Fig. 4(f)]. In shallow gradients ($h = 100$) the global maximum in velocity is achieved at the largest cluster size considered, $N = 50$, because larger clusters are more extended in the $-x$ direction [Fig. 4(g)]. This differs from the sharp-gradient result [Fig. 4(c)]. In shallow gradients, increasingly large clusters gain access to the cells that are best positioned to sense the gradient—those at smaller x (Fig. 2).

Experiments often track gradient-sensing responses as a function of the cluster size [1,3,39,42,43]. Our results show that, without specifying how leaders are chosen, even the qualitative dependence of chemotactic velocity or correlation time on cluster size is not known. We show how $\langle V_x \rangle$ and T_c depend on cluster size for $h = 1$ with three leader allocation strategies (100% leaders, 40% leaders, and choosing the number of leaders that maximizes $\langle V_x \rangle$) in Figs. 4(c) and 4(d). When all cells are leaders, $\langle V_x \rangle$ monotonically decreases while T_c remains constant. If a fixed fraction, 2/5, of cells

are leaders, $\langle V_x \rangle$ first increases, then decreases in cluster size, while T_c increases. And if we choose the leaders to maximize $\langle V_x \rangle$, we see that larger clusters do not slow much—but they do see a steep increase in T_c. These effects depend on the gradient shape, and the situation is much different for wider gradients, where increasing cluster size can increase $\langle V_x \rangle$ or even decrease T_c, as in Fig. 4(h).

The correlation time can also be altered by follower noise (Appendix E). We find that increasing follower noise can decrease the long correlation times at small leader fractions, but that even for high levels of follower noise, there is still a significant increase in correlation time as the number of leaders is decreased.

Though our results so far are for linear clusters, the most critical qualitative features [e.g., Fig. 3(c)] are consistent between compact and linear clusters (Appendix F). The deviations between compact and extended clusters are easily understood in terms of the curves for Δ_ℓ. For instance, in sharp transitions ($h = 1$), compact clusters have a higher $N_\ell^{\max}$, because there are more cells close to the transition $x = 0$.

We have also assumed that the only source of leader error in gradient sensing is ligand-receptor binding. Intracellular noise may also be significant at sharp enough gradients [27]. We study adding a downstream intracellular noise in quadrature, $\Delta_\ell^2 = \Delta_\phi^2 + \Delta_{\text{int}}^2$. The qualitative dependence of $N_\ell^{\max}$ does not significantly change for low ($\sim 4^\circ$) or moderate ($\sim 18^\circ$) levels of downstream noise but becomes substantially washed out at high ($\sim 36^\circ$) levels (Appendix G). This is expected: if the primary source of measurement error does not depend on the environment, specialization to leaders and followers will not be environmentally dependent.

Our results show that choosing whether to specialize to leader and follower cells in a chemotaxing cluster is subtle. Clusters can chemotax more quickly if some cells sense and some cells follow, but the number of leaders to maximize directed migration depends heavily on the gradient and the accuracy of follower cells. In addition, cells at the front of the cluster may not be the most informed. Cells near the middle or back may be provide more accurate directional cues, consistent with experiments indicating that collectives are not necessarily steered by the cells at their front [12]. Our work shows that rear-steering might be optimal in shallow gradients. Cluster persistence times are dramatically increased by leader specialization, and different strategies of allocating leaders trade off chemotactic speed with a cluster's ability to reorient.

Experiments on the zebrafish lateral line have tested to what extent chemosening cells can rescue the migration of clusters with reduced chemosensing ability [7,8]. References [16] and [9] found that cells within 120 μm ($\sim$10 cell diameters [44]) of the front of the primordium experience a concentration gradient. Reference [7] was able to restore migration of a chemosensing reduced primordium with only a few fully chemosensing cells near the front. However, the recent experiments in [8] have shown that the chemosensing ability of cells behind the tip of the primordium also affects its migration. These experiments are consistent with our notion that cell clusters are not steered just by the cells at their leading edge.

We have studied the behavior of cell clusters when we vary the transition width, follower noise level, and cluster size. However, because we are considering a self-generated gradient, it is possible that the cluster velocity or size affects the concentration profile. In past work on the zebrafish lateral line, it was found that diffusion was the primary determinant of the gradient shape [16], so in some cases it may be reasonable to consider the value of h in our model to be fixed by the ligand diffusion coefficient D and the cluster degradation rate k_-, $h \sim \sqrt{D/k_-}$. Even if this is not a good approximation in other cases—that is, the cluster size does affect the gradient shape—the steady-state behavior can be analyzed with the length scale $h(N)$ that emerges given the cluster size.

We have intentionally chosen a highly simplified model for cell-cell interactions. In a real cell cluster, there may be a finite length scale λ over which correlations between cells decay [45]. This might affect the correlation times of large clusters, where cells are separated beyond the scale of λ. Even for clusters larger than λ, our results will provide intuition about how to specialize within subclusters of size λ within the full cluster. Other modifications, like resolving detailed forces between cells or modifying the alignment term so that cells align with their nearest neighbors instead of the collective velocity, might also alter the results. However, the nonmonotonic dependence of $N_\ell^{\max}$ as a function of the transition width in Fig. 3(b) occurs in both the full model and the independent follower model, and for multiple cluster shapes, indicating that this result has some robustness to model details. Future experiments varying the chemosensing ability of cells at different positions within a cluster will be an important way to test the leader-follower mechanism and whether the cells best able to sense the gradient are the ones that drive the cluster's directionality.

ACKNOWLEDGMENTS

We thank Holger Knaut for useful conversations and Alma Plaza-Rodriguez for performing initial simulations on a related model. We acknowledge funding from an IDIES Seed Grant and Johns Hopkins University. This research project was conducted using computational resources at the Maryland Advanced Research Computing Center (MARCC).

APPENDIX A: EXTENDING THE HU *ET AL.* RESULT BEYOND THE SHALLOW GRADIENT APPROXIMATION

As in [22] and [23], we consider a circular cell with N_r receptors uniformly spaced along its perimeter. Let ϕ denote the direction of the gradient. We model the receptors on the cell $x_1 \ldots x_{N_r}$ as N_r independent Bernoulli trials that can be occupied with value 1 or unoccupied with value 0. For simple ligand-receptor kinetics, the probability of receptor n's being occupied given concentration C_n at the receptor and ligand-receptor dissociation constant K_d is $\frac{C_n}{C_n+K_d}$ and the probability of being unoccupied is $\frac{K_d}{C_n+K_d}$. The probability that the nth receptor is occupied is a function of the gradient direction ϕ. Thus, the probability distribution function for the nth receptor is

$$f_{n,\phi}(x_n) = \left(\frac{C_n}{C_n + K_d}\right)^{x_n} \left(\frac{K_d}{C_n + K_d}\right)^{1-x_n}. \quad \text{(A1)}$$

The likelihood function for a cell estimating the gradient direction ϕ given the values at the receptors is

$$\mathcal{L}(\phi|x_1 \ldots x_n) = \prod_1^{N_r} f_{n,\phi}(x_n). \tag{A2}$$

The log-likelihood function is

$$\ln\mathcal{L} = \sum_n \left[x_n \ln\frac{C_n}{C_n + K_d} + (1 - x_n)\ln\frac{K_d}{C_n + K_d} \right] \tag{A3}$$

$$= \sum_n \left[x_n \ln\frac{C_n}{K_d} + \ln\frac{K_d}{C_n + K_d} \right] \tag{A4}$$

$$= \sum_n [x_n(\ln C_n - \ln K_d) + \ln K_d - \ln(C_n + K_d)]. \tag{A5}$$

For an estimation of the gradient direction, we are interested in computing the second derivative of the log-likelihood function with respect to ϕ. Therefore, let a prime denote a derivative taken with respect to ϕ. Then the first derivative of the log-likelihood function with respect to ϕ is

$$\frac{\partial}{\partial\phi}\ln\mathcal{L} = \sum_n \left[x_n \frac{C_n'}{C_n} - \frac{C_n'}{C_n + K_d} \right]. \tag{A6}$$

Taking another derivative gives

$$\frac{\partial^2}{\partial\phi^2}\ln\mathcal{L} = \sum_n \left[x_n \frac{C_n C_n'' - (C_n')^2}{C_n^2} - \frac{(C_n + K_d)C_n'' - (C_n')^2}{(C_n + K_d)^2} \right]. \tag{A7}$$

To find the expectation value, first note that for the Bernoulli trials x_n, the expectation value is just the probability of occupancy $\langle x_n \rangle = \frac{C_n}{C_n + K_d}$. With this expression, the expectation value of the second derivative becomes

$$\left\langle \frac{\partial^2}{\partial\phi^2}\ln\mathcal{L} \right\rangle = \sum_n \left[\langle x_n \rangle \frac{C_n C_n'' - (C_n')^2}{C_n^2} - \frac{(C_n + K_d)C_n'' - (C_n')^2}{(C_n + K_d)^2} \right] \tag{A8}$$

$$= \sum_n \left[\frac{C_n}{C_n + K_d} \frac{C_n C_n'' - (C_n')^2}{C_n^2} - \frac{(C_n + K_d)C_n'' - (C_n')^2}{(C_n + K_d)^2} \right] \tag{A9}$$

$$= \sum_n -\left[\frac{K_d (C_n')^2}{C_n (C_n + K_d)^2} \right]. \tag{A10}$$

This expression holds true for any gradient profile and does not make any assumptions about its steepness.

In this work, the goal is to compute the uncertainty for leader cells with the gradient varying the in x direction given by $C(x) = \frac{1}{2}C_{\max}(1 + \frac{1}{2}\tanh(\frac{x}{h}))$. For the ith cell in the cluster, the x position of the center is $1 - i$, where i ranges from 1 to the cluster size N. Since we are working in units in which the cell diameter is 1, the x positions of the receptors are given by

$$x(i, \alpha_n) = 1 - i + \tfrac{1}{2}\cos(\alpha_n - \phi), \tag{A11}$$

where α_n is the angular position of the nth receptor. The sum in Eq. (A10) can be approximated as an integral for a large number of receptors. Therefore, for a given cell i, the Fisher information is

$$\mathcal{I}(i) = -\left\langle \frac{\partial^2}{\partial\phi^2}\ln\mathcal{L}(i) \right\rangle \approx \frac{N_r}{2\pi}\int_0^{2\pi} \frac{K_d (C_n'(\alpha, i))^2}{C_n(\alpha, i)(C_n(\alpha, i) + K_d)^2} d\alpha, \tag{A12}$$

where the concentration and its derivative are written in terms of α and i using Eq. (A11). This expression does not depend on the gradient direction ϕ because the receptors are evenly spaced, and the cell is symmetric under rotations in the continuum limit. We numerically integrate Eq. (A12) with Gaussian quadrature to determine the Fisher information for each cell in the cluster. With the Fisher information, the Cramér-Rao bound gives the minimum uncertainty for the gradient direction estimated by the ith cell as

$$\Delta_\phi(i) = (\mathcal{I}(i))^{-1/2}. \tag{A13}$$

APPENDIX B: DETERMINING σ AND τ AS A FUNCTION OF Δ

We want Eqs. (3) and (4) to represent cells with given angular uncertainties Δ_i. How can we map between Δ and the parameters of the model for each cell, τ and σ? To answer this question, we have to understand a bit more about the solutions of these equations, which are of the form

$$\frac{d\theta(t)}{dt} = -\frac{\theta(t)}{\tau} + \sigma\xi(t) \quad \text{for} \quad \theta \in [-\pi, \pi]. \tag{B1}$$

Without periodic boundary conditions, the dynamics of θ would follow an Ornstein-Uhlenbeck process [46], which can be described in terms of a known steady-state distribution [normal with mean 0 and variance $\frac{\sigma^2\tau}{2}$, as in Eq. (B4)] and a transition probability [Eq. (B6)], which is the distribution of θ at a time t' given the value of θ at an earlier time t. However, the periodic boundary conditions mean that the dynamics of θ are more complicated than a standard Ornstein-Uhlenbeck process. The steady-state distribution becomes a truncated normal distribution determined from renormalizing Eq. (B4) in an interval $[-\pi, \pi]$. This results in the steady-state probability distribution

$$p(\theta) = \begin{cases} [\sigma\sqrt{\pi\tau}(\operatorname{erf}(\sqrt{2}\pi/\sigma^2\tau))]^{-1}\exp\left(-\frac{\theta^2}{\sigma^2\tau}\right), & \theta \in [-\pi, \pi], \\ 0 & \text{otherwise.} \end{cases} \tag{B2}$$

On the other hand, the transition probability is not analytically tractable. Therefore, we use results from the Ornstein-Uhlenbeck process in the limits of very small and very large noises Δ_i but must use numerical methods for intermediate values of Δ_i.

For each cell i, the fluctuations of its angle θ_i about its mean value are characterized by the term $\frac{\sigma_i^2\tau_i}{2}$ [Eq. (B2)]. Therefore, we set the scale of these fluctuations equal to the angular noise Δ_i through the equation

$$\Delta_i^2 = \frac{\sigma_i^2\tau_i}{2}. \tag{B3}$$

We note here that Δ_i^2 can be larger than 2π; here, we are setting the variance of the *parent normal* that is truncated to find Eq. (B2).

If we applied only our formula for Δ_i [Eq. (B3)], there would not be a unique way to choose both σ_i and τ_i; we need to do more than just fix the variance of the angle $\langle\delta\theta_i^2\rangle$. Therefore, we develop a procedure to choose values for σ_i and τ_i so that the behavior of the cell is realistic at all noise levels. If we naively chose $\sigma_i \sim \Delta_i$, then as Δ_i became large, the cell would undergo angular diffusion with a diverging diffusion coefficient—physically unrealistic. We want to find functions $\sigma(\Delta)$ and $\tau(\Delta)$ so that the correlation time $T_{\mathbf{P}}$ of a single cell's polarity $\mathbf{P}_i$ is equal to 1 at all angular uncertainties. This corresponds to the idea that the cell has a constant time to reorient, independent of how accurately it is measuring its environment.

We first consider a Gaussian approximation, which does not account for the periodicity of the angle θ, to derive the asymptotic expressions for $\sigma(\Delta)$ and $\tau(\Delta)$. Then we numerically determine values for $\sigma(\Delta)$ and $\tau(\Delta)$ at intermediate levels of angular uncertainties.

1. Gaussian approximation

For an angle $\theta(t)$ relaxing to 0 with noise σ and time constant τ, the equation governing θ is Eq. (B1). However, in the Gaussian approximation, the periodicity of the variable θ is ignored. Although this result is not generally applicable, it is asymptotically correct in the limits of very small and very large angular noise. This is an Ornstein-Uhlenbeck process whose steady-state distribution is a normal distribution with mean 0 and variance $\frac{\sigma^2\tau}{2}$:

$$\theta(t) \sim N\left(0, \frac{\sigma^2\tau}{2}\right). \tag{B4}$$

An angle at a later time $\theta(t+t')$, given the value of $\theta(t)$, will have the distribution

$$\theta(t+t') \sim N\left(\theta(t)e^{-t'/\tau}, \frac{\sigma^2\tau}{2}(1-e^{-2t/\tau})\right). \tag{B5}$$

The first equation is the marginal distribution of $\theta(t)$ and the second equation is the conditional distribution of $\theta(t+t')$ given $\theta(t)$, so the joint probability distribution $p(\theta(t+t');\theta(t))$ is just the product of the two distributions. Let $\theta(t) = x_1$ and $\theta(t+t') = x_2$ for ease of notation. Then the joint probability density function is

$$f(x_1,x_2) = \frac{1}{\pi\sigma^2\tau\sqrt{1-e^{-2t'/\tau}}} \times e^{-x_1^2/\sigma^2\tau} e^{-(x_2-x_1e^{-t'/\tau})^2/\sigma^2\tau(1-e^{-2t'/\tau})}. \tag{B6}$$

For a polarity vector $\mathbf{P}(t) = (\cos\theta(t), \sin\theta(t))$, the time correlation function is given by

$$\phi(t') = (\langle\mathbf{P}(t)\cdot\mathbf{P}(t+t')\rangle - \langle\mathbf{P}(t)\rangle \cdot \langle\mathbf{P}(t+t')\rangle)/(\langle\mathbf{P}(t)\cdot\mathbf{P}(t)\rangle - \langle\mathbf{P}(t)\rangle^2), \tag{B7}$$

which at steady state depends only on the separation in time and is normalized so that $\phi(0) = 1$.

The term $\langle\mathbf{P}(t)\cdot\mathbf{P}(t+t')\rangle$ is given by the following expression:

$$\langle\mathbf{P}(t)\cdot\mathbf{P}(t+t')\rangle = \int_{-\infty}^{\infty} dx_1dx_2 f(x_1,x_2)[\cos(x_1)\cos(x_2) + \sin(x_1)\sin(x_2)] \tag{B8}$$

$$= \int_{-\infty}^{\infty} dx_1dx_2 f(x_1,x_2)\cos(x_2-x_1). \tag{B9}$$

To evaluate the this term, we change the variables in the joint density function in Eq. (B6) through the following transformation:

$$y_1 = x_1,$$
$$y_2 = x_2 - x_1e^{-t'/\tau}. \tag{B10}$$

The Jacobian from this transformation is 1, so the new joint probability distribution is

$$g(y_1,y_2) = \frac{1}{\pi\sigma^2\tau\sqrt{1-e^{-2t'/\tau}}} e^{-y_1^2/\sigma\tau} e^{-y_2^2/\sigma^2\tau(1-e^{-2t'/\tau})}. \tag{B11}$$

We can apply this transformation to Eq. (B9) and use the distribution in Eq. (B11) to evaluate the expectation value in the following way:

$$\langle\mathbf{P}(t)\cdot\mathbf{P}(t+t')\rangle = \int_{-\infty}^{\infty} dx_1dx_2 f(x_1,x_2)\cos(x_2-x_1) \tag{B12}$$

$$= \int_{-\infty}^{\infty} dy_1dy_2 g(y_1,y_2)\cos(y_2+y_1(e^{-t'/\tau}-1)) \tag{B13}$$

$$= \int_{-\infty}^{\infty} dy_1dy_2 \frac{1}{\pi\sigma^2\tau\sqrt{1-e^{-2t'/\tau}}} e^{-y_1^2/\sigma\tau} \times e^{-y_2^2/\sigma^2\tau(1-e^{-2t'/\tau})}\cos(y_2+y_1(e^{-t'/\tau}-1)). \tag{B14}$$

Applying the integral

$$\frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\cos(ax+b)e^{-x^2/2\sigma^2}dx = \cos(b)e^{-a^2\sigma^2/2} \tag{B15}$$

twice (once for y_1 and once for y_2) to Eq. (B14) gives the result

$$\langle\mathbf{P}(t)\cdot\mathbf{P}(t+t')\rangle = \exp\left(\frac{\sigma^2\tau}{2}(-1+e^{-t'/\tau})\right). \tag{B16}$$

At steady state, the mean polarities will be independent of time, and the identity

$$\langle \mathbf{P}(t)\rangle = \langle \mathbf{P}(t+t')\rangle = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\sigma^2\tau}}\cos(\theta)e^{-\theta^2/\sigma^2\tau}d\theta\hat{x} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi\sigma^2\tau}}\sin(\theta)e^{-\theta^2/\sigma^2\tau}d\theta\hat{y} \tag{B17}$$

$$= \exp(-\sigma^2\tau/4)\hat{x} + 0\hat{y} \tag{B18}$$

holds from applying Eq. (B15). Therefore, the term $\langle \mathbf{P}(t)\rangle \cdot \langle \mathbf{P}(t+t')\rangle$ evaluates as

$$\langle \mathbf{P}(t)\rangle \cdot \langle \mathbf{P}(t+t')\rangle = \exp(-\sigma^2\tau/2). \tag{B19}$$

Thus, the final expression for $\phi(t')$, normalized by its value at $t'=0$, is

$$\phi(t') = \frac{\exp\left(\frac{\sigma^2\tau}{2}(-1+e^{-t'/\tau})\right) - \exp(-\sigma^2\tau/2)}{1-\exp(-\sigma^2\tau/2)} = \frac{\exp\left(\frac{\sigma^2\tau}{2}e^{-t'/\tau}\right)-1}{e^{\frac{\sigma^2\tau}{2}}-1}. \tag{B20}$$

The correlation time is defined as the integral of this correlation function from $t'=0$ to $t'=\infty$. Only the numerator of Eq. (B20) needs to be evaluated since the denominator does not depend on t'. Therefore, the integral to evaluate is the integral I:

$$I = \int_0^{\infty}\left(\exp\left(\frac{\sigma^2\tau}{2}e^{-t'/\tau}\right)-1\right)dt'. \tag{B21}$$

With the substitution $u = \frac{\sigma^2\tau}{2}e^{-t'/\tau}$, $du = -\frac{1}{\tau}udt'$, the expression for the integral becomes

$$I = \tau\int_0^{\frac{\sigma^2\tau}{2}}\frac{(e^u-1)}{u}du. \tag{B22}$$

This integral can be solved and gives

$$I = \mathrm{Ei}\left(\frac{\sigma^2\tau}{2}\right) - \gamma - \log\left(\frac{\sigma^2\tau}{2}\right), \tag{B23}$$

where γ is the Euler-Mascheroni constant and the exponential integral Ei is defined as

$$\mathrm{Ei}(x) = \int_{-\infty}^{x}\frac{e^t}{t}dt. \tag{B24}$$

Thus, the correlation time, $T_{\mathbf{P}}$, of the polarity $\mathbf{P}$ is

$$T_{\mathbf{P}} = \int_0^{\infty}\phi(t')dt' = \frac{I}{\exp\left(\frac{\sigma^2\tau}{2}\right)-1} = \tau\frac{-\gamma+\mathrm{Ei}\left(\frac{\sigma^2\tau}{2}\right)-\log\left(\frac{\sigma^2\tau}{2}\right)}{\exp\left(\frac{\sigma^2\tau}{2}\right)-1} \tag{B25}$$

in terms of the angular relaxation time τ and angular noise σ.

To ensure that at every level of angular noise the polarity correlation time is constant, the polarity correlation time is set equal to 1 and the angular relaxation time τ and angular noise σ are chosen so that the angular uncertainty is as required. This can be described by the equations

$$T_{\mathbf{P}} = 1, \tag{B26}$$

$$\frac{\sigma^2\tau}{2} = \langle\delta\theta^2\rangle = \Delta_\phi^2, \tag{B27}$$

so that once a cell's angular uncertainty Δ_ϕ^2 is known, the correct τ and σ for that cell can be chosen to set its polarity correlation time equal to 1. Thus, the formulas for $\tau(\Delta_\phi)$ and $\sigma(\Delta_\phi)$ are

$$\tau(\Delta_\phi) = \frac{\exp\left(\Delta_\phi^2\right)-1}{-\gamma+\mathrm{Ei}\left(\Delta_\phi^2\right)-\log\left(\Delta_\phi^2\right)}, \tag{B28}$$

$$\sigma(\Delta_\phi) = \Delta_\phi\sqrt{\frac{2}{\tau(\Delta_\phi)}}. \tag{B29}$$

However, these formulas are only asymptotically correct because they do not account for periodic boundary conditions. Thus, the two limits we use give the following equations for the asymptotic forms:

$$\tau(\Delta_\phi)\approx 1, \quad \sigma(\Delta_\phi)\approx\sqrt{2}\Delta_\phi \quad \text{for } \Delta_\phi\to 0, \tag{B30}$$

$$\tau(\Delta_\phi)\approx\Delta_\phi^2, \quad \sigma(\Delta_\phi)\approx\sqrt{2} \quad \text{for } \Delta_\phi\to\infty. \tag{B31}$$

2. Numerical interpolation

For intermediate values of Δ_ϕ^2, the parameters τ and σ determined from the Gaussian approximation can give real correlation times that deviate by up to 25% from the desired value of 1. This deviation is due to the wrapping of θ in the interval $[-\pi,\pi]$, which is not accounted for in the Gaussian approximation. To accurately incorporate the effects of periodic boundary conditions, we want to numerically find values for the parameters $\tau(\Delta_\phi)$ and $\sigma(\Delta_\phi)$ so that the relations $T_{\mathbf{P}}=1$ and $\Delta_\phi^2=\frac{\sigma^2\tau}{2}$ are both true. We simulate an angle following Eq. (B1). We ensure that θ remains in the interval $[-\pi,\pi]$ by adding π, computing θ modulo 2π, then shifting θ back to the interval $[-\pi,\pi]$ by subtracting π. That is, at each time step, we apply the following formula:

$$\theta_\ell = ((\theta_\ell+\pi)\bmod 2\pi)-\pi. \tag{B32}$$

In the Gaussian approximation, the quantity $\Delta_\phi^2=\frac{\sigma^2\tau}{2}$ is the variance in the steady-state Gaussian distribution of θ. With periodic boundaries, the steady-state distribution of θ is a truncated normal distribution in $[-\pi,\pi]$. Therefore, $\Delta_\phi^2=\frac{\sigma^2\tau}{2}$ is the variance of the parent normal distribution to the truncated normal distribution for θ [as in Eq. (B2)]. Thus, small values of Δ_ϕ^2 result in an approximately Gaussian distribution of θ around the gradient direction, while very large values of Δ_ϕ^2 correspond to a uniformly distributed value of θ, corresponding to a cell that chooses its direction randomly. However, the transition probability from one angle to another is not analytically tractable with periodic boundaries, so we need simulations to determine the correlation time $T_{\mathbf{P}}$.

As implied by the form of Eq. (B25) and the Buckingham Pi theorem, the equation for the correlation time can be written in terms of the parameters τ and σ as

$$T_{\mathbf{P}}\left(\Delta_\phi^2\right) = \tau f\left(\frac{\sigma^2\tau}{2}\right). \tag{B33}$$

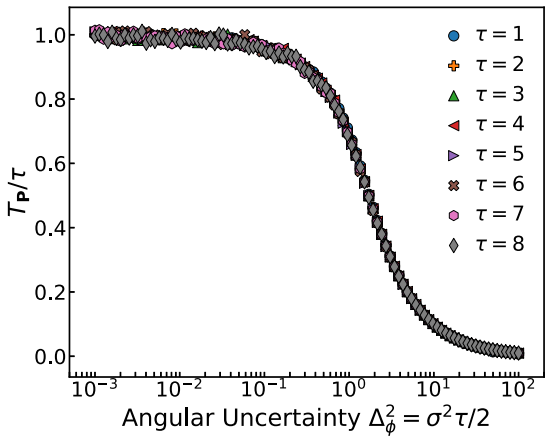

FIG. 5. Universality of the single-cell polarity correlation time scaling. The scaling between the correlation time of the polarity $T(\Delta_\phi^2)$ normalized by the parameter τ is universal for all values of τ. Here, τ is chosen as a parameter and $\sigma = \sqrt{2}\Delta_\phi/\sqrt{\tau}$.

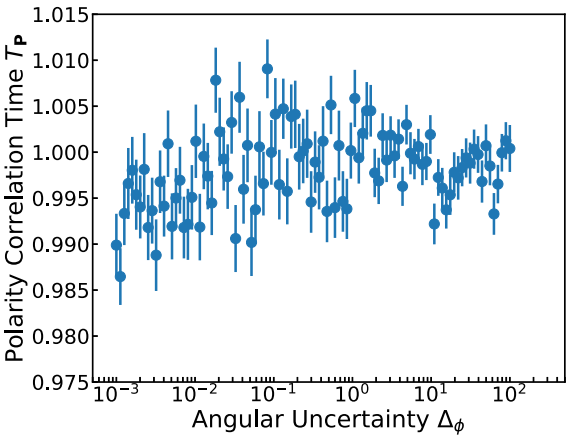

FIG. 6. Simulated tests of the numerical scheme. The correlation time of the polarity T_P for a single cell relaxing to an angle of 0 is plotted as a function of its uncertainty Δ_ϕ. The parameters $\sigma(\Delta_\phi)$ and $\tau(\Delta_\phi)$ are chosen according to the process described above. The correlation time remains within 2% of the desired time of 1.

This is also shown empirically in Fig. 5, where for various values of τ and Δ_ϕ^2 the polarity correlation time $T_P(\Delta_\phi^2)$ is determined from simulations with $\sigma = \sqrt{2}\Delta_\phi/\sqrt{\tau}$. Therefore, without loss of generality, we choose the parameter $\tau = 1$ and $\sigma = \sqrt{2}\Delta_\phi$ and simulate $T_P(\Delta_\phi)$ to find the function $f(\frac{\sigma^2\tau}{2})$. This function is useful because choosing $\tau(\Delta_\phi) = 1/f(\frac{\sigma^2\tau}{2})$ ensures that the correlation time of the polarity $T_P = 1$ at all values of Δ_ϕ. Thus, we choose the parameters according to the rule

$$\tau(\Delta_\phi) = 1/f\left(\frac{\sigma^2\tau}{2}\right), \tag{B34}$$

$$\sigma(\Delta_\phi) = \Delta_\phi\sqrt{2f\left(\frac{\sigma^2\tau}{2}\right)}. \tag{B35}$$

To find the correlation time, we simulate an angle relaxing to 0 with periodic boundaries enforced (see Appendix C). We use a time step of $\Delta t = 0.01$, as in the main simulations, and we simulate to a time of 3000, which was found to generate good statistics for the correlation function. The correlation time is found by fitting an exponential to the correlation function, though rescaling the time at which the correlation function reaches $1/2$ to determine when a $1/e$ decay would have occurred gives similar trends. We repeat this procedure to generate 100 measurements of T_P for each value of Δ_ϕ^2. Then the parameters $\tau(\Delta_\phi)$ and are found through Eqs. (B34) and (B35) and linear interpolation. A grid is created for $0.01 \leqslant \Delta_\phi^2 \leqslant 10$, where, for each range of orders of magnitude, the grid spacing is $1/10$ of the smallest value. Outside those limits, we use the asymptotic forms derived using the Gaussian approximation.

3. Verifying numerical scheme

To determine the accuracy of the numerical method, we compute T_P for single cells of various angular uncertainties Δ_ϕ, where we choose the parameters τ and σ according to the procedure outlined in the previous section. In Fig. 6, we present values computed for angular uncertainty levels that lie in both asymptotic limits. The computed values are within at least 2% of the desired time of 1. Values outside the range considered here will be at least this close to 1 because the asymptotic limits will improve at very small or very large Δ_ϕ.

APPENDIX C: SIMULATION DETAILS

Simulation code is provided at Ref. [47]. The cluster velocities are simulated until $t = 5200$ for various gradient widths, cluster sizes, follower noises, and number of leaders. Steady state is considered to be reached after $t = 200$, since that is approximately the longest correlation time encountered. From the steady-state data, we compute the correlation function $\varphi(t') = \langle \mathbf{V}_c(t) \cdot \mathbf{V}_c(t+t')\rangle - \langle \mathbf{V}_c(t)\rangle^2$ and determine a correlation time by fitting an exponential function to $\varphi(t')/\varphi(0)$. We show a representative example in Fig. 7, which is the

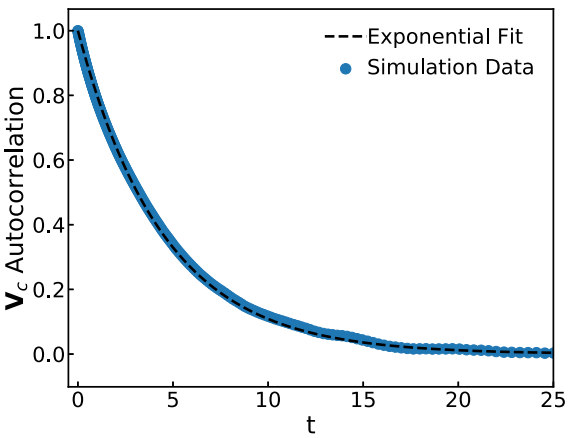

FIG. 7. Representative exponential fit for the autocorrelation function. The correlation function and an exponential fit are compared for a train of $N = 50$ cells in an $h = 10$ gradient with 10 leader cells and followers with $\Delta_f = 36^\circ$. We use the exponential fit to determine the correlation time.

steady-state cluster velocity autocorrelation and its exponential fit for a train of $N = 50$ cells in an $h = 10$ gradient with 10 leader cells and followers with $\Delta_f = 36^\circ$. Rescaling the time at which the correlation function reaches 1/2 to determine when a 1/e decay would have occurred gives similar trends. After the correlation time T_c for the cluster has been measured for a run, the steady-state velocity data are broken up into intervals of $3T_c$ so that the mean of each interval is an independent measurement of the steady-state velocity. We repeat the simulation 50 times to generate 50 measurements of the correlation time and many measurements of the steady-state velocity. The 95% confidence intervals are generated for the mean velocity in the gradient direction $\langle V_x \rangle$ and the correlation time T_c from the 50 samples using bootstrap methods [48]. To determine the error bars on $N_\ell^{\max}$, we use the distributions of $\langle V_x \rangle$ generated from the bootstrap procedure. For fixed gradient and cluster properties, we draw a value of $\langle V_x \rangle(N_\ell)$ for each possible value of N_ℓ. Then for each draw we record which value of N_ℓ gives the highest $\langle V_x \rangle$. We draw 10 000 times to generate a distribution for $N_\ell^{\max}$. This procedure produces a distribution of the number of leaders which optimizes a given quantity.

We account for periodic boundary conditions for both the leader and the follower cells. For follower cells, we compute $\theta_f - \theta_c$ with the arctan2 function as

$$\theta_f - \theta_c = \text{arctan2}(V_x P_{f,y} - V_y P_{f,x}, V_x P_{f,x} + V_y P_{f,y}), \quad \text{(C1)}$$

where the first term is the cross product $\mathbf{V}_c \times \mathbf{P}_f = |\mathbf{V}_c||\mathbf{P}_f|\sin(\theta_f - \theta_c)$, the second is the dot product $\mathbf{V}_c \cdot \mathbf{P}_f = |\mathbf{V}_c||\mathbf{P}_f|\cos(\theta_f - \theta_c)$, and arctan2($b$,$a$) returns the angle of a vector $\mathbf{v} = a\hat{x} + b\hat{y}$. Therefore, Eq. (C1) will return $\theta_f - \theta_c$, and the function is defined so that the angle is in the interval $[-\pi, \pi]$. For leader cells, we apply Eq. (B32) to their angles at each time step. This is equivalent to the procedure in Eq. (C1) if $V_x = 1$ and $V_y = 0$; in either case, the angle relative to the x axis wrapped in $[-\pi, \pi]$ is returned.

To avoid numerical errors associated with division by 0, the minimum of the concentration is taken as 10^{-14} to avoid dividing by 0 when computing the leader cell uncertainty for cells far from the transition region. Since that is a value much lower than the dissociation constant, the leaders are just as random as they would be if the value were truly 0, and this does not impact the results.

APPENDIX D: INDEPENDENT FOLLOWER APPROXIMATION

The cluster velocity is given by the mean polarity of all the cells in the cluster, which can be decomposed into leader and follower contributions:

$$\mathbf{V}_c = \frac{1}{N}\sum_i \mathbf{P}_i = \frac{1}{N}\left(\sum_\ell \mathbf{P}_\ell + \sum_f \mathbf{P}_f\right). \quad \text{(D1)}$$

Therefore, the mean x velocity of the cluster is given by

$$\begin{aligned}\langle V_c^x \rangle &= \frac{1}{N}\left(\sum_\ell \langle P_\ell^x \rangle + \sum_f \langle P_f^x \rangle\right) \\ &= \frac{1}{N}\left(\sum_\ell \langle \cos(\theta_\ell) \rangle + \sum_f \langle \cos(\theta_f) \rangle\right) \quad \text{(D2)}\end{aligned}$$

in terms of the leader angle θ_ℓ and follower angle θ_f. The followers relax towards the angle of the cluster θ_c with some noise. The angle of a follower θ_f can be written in terms of the angle of the cluster and some angle $\theta_{\text{rel},f}$, which is the angle of follower f relative to the cluster angle

$$\theta_f = \theta_c + \theta_{\text{rel},f}. \quad \text{(D3)}$$

The x polarity of a follower is just the cosine of the follower angle, which can be rewritten using Eq. (D3) as

$$\begin{aligned}\cos(\theta_f) &= \cos(\theta_c + \theta_{\text{rel},f}) \\ &= \cos(\theta_c)\cos(\theta_{\text{rel},f}) + \sin(\theta_c)\sin(\theta_{\text{rel},f}), \quad \text{(D4)}\end{aligned}$$

and the mean is given by

$$\langle \cos(\theta_f) \rangle = \langle \cos(\theta_c)\cos(\theta_{\text{rel},f}) \rangle + \langle \sin(\theta_c)\sin(\theta_{\text{rel},f}) \rangle. \quad \text{(D5)}$$

So far, the above results are exact. To develop the approximation, we make two key assumptions. First, we assume that the relative angle of the follower $\theta_{\text{rel},f}$ is independent of the cluster angle θ_c; this can be simplified as

$$\langle \cos(\theta_f) \rangle = \langle \cos(\theta_c) \rangle \langle \cos(\theta_{\text{rel},f}) \rangle \quad \text{(D6)}$$

because θ_c is symmetric about 0 so $\langle \sin(\theta_c) \rangle = 0$. To evaluate the expectation value $\langle \cos(\theta_{\text{rel},f}) \rangle$, we note that its dynamics

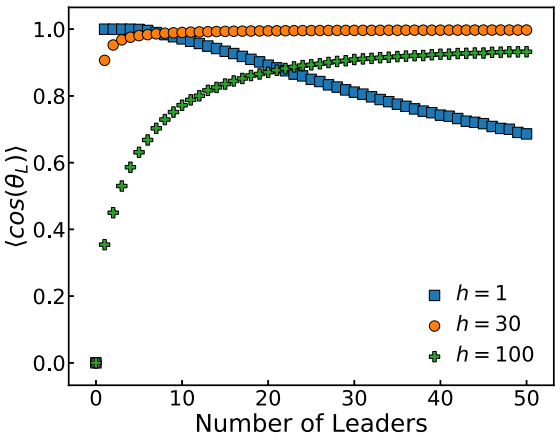

FIG. 8. Leader signals in the independent follower approximation. In the independent follower approximation, the signal to the followers is the cosine of the polarity vector of all the leaders. For sharp gradients ($h \sim 1$), the first few leaders are well informed and the cosine is near 1, but it decreases as essentially random leaders are added. For intermediate gradients ($h \sim 30$), the curve is flat and large, as all the leaders have similar and low uncertainties in the gradient direction. For wide gradients ($h \sim 100$), the direction of the leaders becomes more accurate as more are added, with diminishing returns on additional leaders.

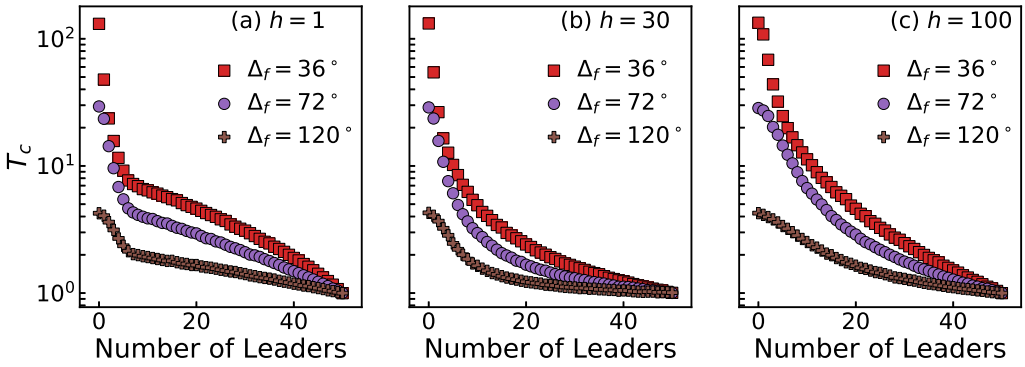

FIG. 9. The cluster correlation time decreases with increasing follower noise. The cluster correlation time T_c decreases as a function of the follower noise Δ_f. Although the extent of the difference in cluster correlation time between the few-leader and all-leader cases quantitatively depends on the follower noise, the trend does not. Shown are the gradients $h = 1$ (a), $h = 30$ (b), and $h = 100$ (c).

are essentially that of the follower angle θ_f in a reference frame in which $\theta_c = 0$. Therefore, the distribution of $\theta_{\mathrm{rel},f}$ will be a truncated normal distribution in $[-\pi, \pi]$ with mean 0 and whose parent normal has a standard deviation Δ_f. From this result, an exact expression for $\langle\cos(\theta_{\mathrm{rel},f})\rangle$ can be applied to find

$$\langle\cos(\theta_f)\rangle = \langle\cos(\theta_c)\rangle\langle\cos(\theta_{\mathrm{rel},f})\rangle \tag{D7}$$

$$= \langle\cos(\theta_c)\rangle \int_{-\pi}^{\pi} \frac{\cos(\theta)}{\sigma\sqrt{\pi\tau}(\mathrm{erf}(\pi/\sqrt{2}))} \times \exp\left(-\theta^2/2\Delta_f^2\right)d\theta \tag{D8}$$

$$= \langle\cos(\theta_c)\rangle e^{-\Delta_f^2/2} \times \mathrm{Re}\left(\mathrm{erf}\left(\left(\pi + i\Delta_f^2\right)/\sqrt{2}\Delta_f\right)\right)/\mathrm{erf}(\pi/\sqrt{2}\Delta_f). \tag{D9}$$

Let the contribution to the polarity from the leaders be defined as the total leader polarity vector $\mathbf{P}_L$:

$$\mathbf{P}_L = \frac{1}{N_\ell}\sum_\ell \mathbf{P}_\ell = |\mathbf{P}_L|(\cos(\theta_L), \sin(\theta_L)). \tag{D10}$$

Second, we assume that the average x component of the cluster is approximated by the average x component of $\mathbf{P}_L$, i.e.,

$$\langle\cos(\theta_c)\rangle \approx \langle\cos(\theta_L)\rangle. \tag{D11}$$

This assumption is generally good since the follower polarities align with the leaders in steady-state. However, it breaks down when the follower correlations contribute significantly to the accuracy of the cluster direction.

Explicitly, the term we use to approximate $\cos(\theta_c)$ is

$$\cos(\theta_L) = \frac{\sum_\ell \cos(\theta_\ell)}{\sqrt{\left(\sum_\ell \cos(\theta_\ell)\right)^2 + \left(\sum_\ell \sin(\theta_\ell)\right)^2}}. \tag{D12}$$

We compute $\langle\cos(\theta_L)\rangle$ numerically. We draw the leader angles θ_ℓ from their respective steady-state truncated normal distributions and compute the mean value of Eq. (D12) from the draws. We use 100 000 draws for $h \leqslant 100$ and 500 000 draws for $h = 125$. Once this term has been computed numerically, Eq. (D9) gives the follower contribution P_f^x to the mean cluster velocity $\langle V_x\rangle$, where we approximate the term $\langle\cos(\theta_c)\rangle$ as $\langle\cos(\theta_L)\rangle$. Then the leader contribution P_ℓ^x to $\langle V_x\rangle$ is simply

$$\left\langle P_\ell^x\right\rangle = \langle\cos(\theta_\ell)\rangle \tag{D13}$$

$$= \int_{-\pi}^{\pi} \frac{\cos(\theta)}{\sigma\sqrt{\pi\tau}(\mathrm{erf}(\pi/\sqrt{2}))}\exp\left(-\theta^2/2\Delta_\ell^2\right)d\theta \tag{D14}$$

$$= e^{-\Delta_\ell^2/2}\mathrm{Re}\left(\mathrm{erf}\left(\left(\pi + i\Delta_\ell^2\right)/\sqrt{2}\Delta_\ell\right)\right)/\mathrm{erf}(\pi/\sqrt{2}\Delta_\ell) \tag{D15}$$

because the leaders independently align with the gradient direction. Thus, the final expression for the gradient velocity is

$$\langle V_x\rangle = \frac{1}{N}\left(\sum_\ell \left\langle P_\ell^x\right\rangle + \sum_f \left\langle P_f^x\right\rangle\right) \tag{D16}$$

$$= \frac{1}{N}\left[\sum_\ell e^{-\Delta_\ell^2/2}\mathrm{Re}\left(\mathrm{erf}\left(\left(\pi + i\Delta_\ell^2\right)/\sqrt{2}\Delta_\ell\right)\right)/\mathrm{erf}(\pi/\sqrt{2}\Delta_\ell) + \sum_f \langle\cos(\theta_L)\rangle e^{-\Delta_f^2/2}\mathrm{Re}\left(\mathrm{erf}\left(\left(\pi + i\Delta_f^2\right)/\sqrt{2}\Delta_f\right)\right)/\mathrm{erf}(\pi/\sqrt{2}\Delta_f)\right], \tag{D17}$$

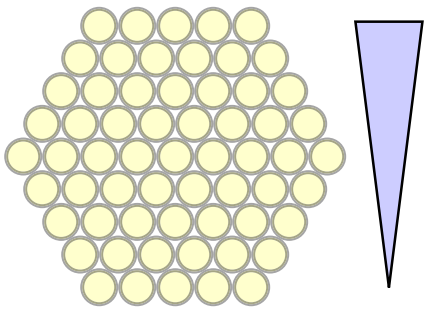

FIG. 10. Illustration of a hexagonally packed cluster with 61 cells.

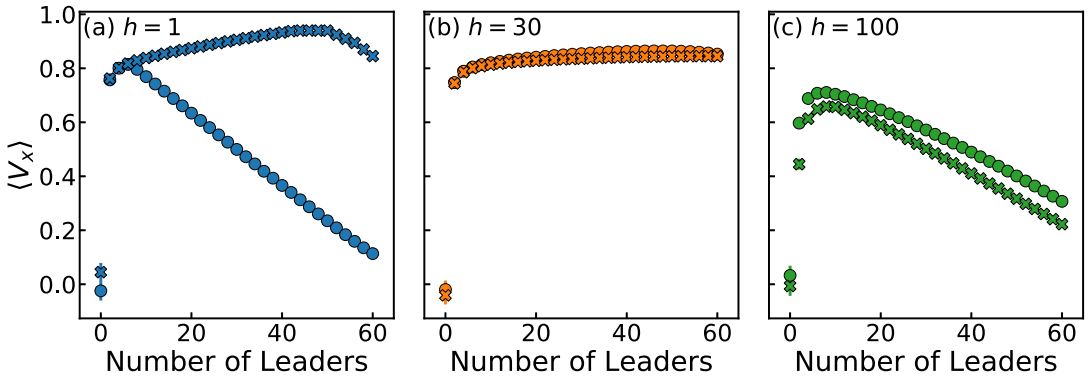

FIG. 11. Cluster velocities for compact and extended clusters. We compare the mean velocity in the gradient direction $\langle V_x \rangle$ between $N = 61$ trains and $Q = 4$ oligomers. The follower noise level is 36°. Circles represent the trains and Xs represent the oligomers. (a) In the sharp $h = 1$ gradient, the oligomer geometry places more cells near the sharp gradient transition, so there are more accurate leaders and the cluster can chemotax more rapidly than the train. (b) The $h = 30$ gradient produces a relatively flat leader uncertainty curve, so the geometry does not have a large impact. (c) The wide $h = 100$ gradient means that the most information about the gradient direction is far from the transition regime, so the larger extent of the train geometry allows it to chemotax slightly more rapidly than the oligomer.

where $\langle \cos(\theta_L) \rangle$ is computed numerically through Eq. (D12), as described above. Once we have $\langle V_x \rangle$ as a function of the leader number, we can compute $N_\ell^{\max}$.

To illustrate the signal that the followers respond to in this approximation, we plot $\cos(\theta_L)$ as leaders are added in Fig. 8. This quantity does not vary as strongly as the x velocity because diffusing leaders do not distort the direction in a consistent way, which mitigates an inaccurate leader's impact on the directionality of $\mathbf{P}_L$.

APPENDIX E: FOLLOWER NOISE AND CORRELATION TIME

In the paper, we show that the cluster correlation time T_c depends strongly on the leader number and the cluster size. However, one might suspect that the correlation time might also depend on the follower noise level Δ_f because the followers maintain the cluster velocity at any instant, driving its persistence. Therefore, at a fixed gradient, we show the dependence of T_c on the leader number for low ($\Delta_f = 36°$), intermediate ($\Delta_f = 72°$), and high ($\Delta_f = 120°$) levels of follower noise in Fig. 9. In the sharp ($h = 1$), medium ($h = 30$), and wide ($h = 100$) gradients, the trend is the same; increasing the follower noise decreases the correlation time at leader fractions less than 1. Also, even at the highest level of follower noise considered here, there is still a significant change in T_c from a small leader number to the all-leader case, and this effect persists at each gradient.

APPENDIX F: CLUSTER GEOMETRY

We compare the train geometry considered in the text to clusters with a compact geometry. We examine the geometry of a four-layer oligomer (as considered in previous work [30,31]), in which 61 cells are hexagonally packed in the cluster, as illustrated in Fig. 10 (we choose the cluster orientation with respect to the gradient as shown in that figure).

To exactly isolate the effects of geometry, we also simulate trains of $N = 61$ cells to compare with the more compact geometry. First, we examine how the geometry affects the mean migration speed, $\langle V_x \rangle$. In Fig. 11, we compare the mean velocity curves as a function of the leader number between trains and oligomers for small ($h = 1$), medium ($h = 30$), and large ($h = 100$) gradient widths at a fixed follower noise level $\Delta_f = 36°$. For the sharp gradient, the hexagonally packed cluster has many more cells near the transition region, so it has many cells that can accurately measure the gradient. Therefore, adding leaders continues to increase its velocity for a larger number of leaders than the $N = 61$ train, and it can migrate more quickly than the train. In the medium gradient width, the accuracy of the leaders does not change

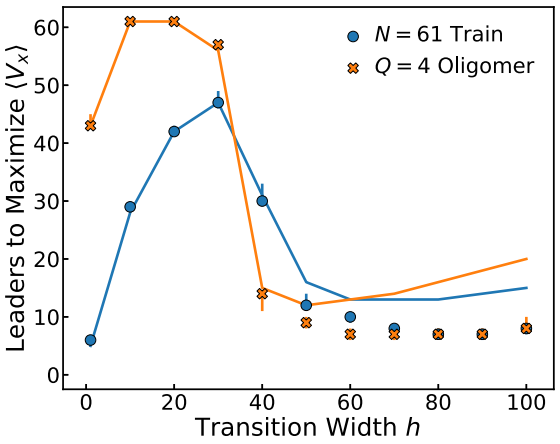

FIG. 12. The number of leaders to maximize the migration speed for compact and extended clusters. We compare $N_\ell^{\max}$ for $N = 61$ trains and $Q = 4$ oligomers. The follower noise level is 36°. Circles represent the trains and Xs represent the oligomers. Lines show the independent follower model. Both geometries follow a similar trend; the number of leaders first increases, then decreases as the gradient transition widens.

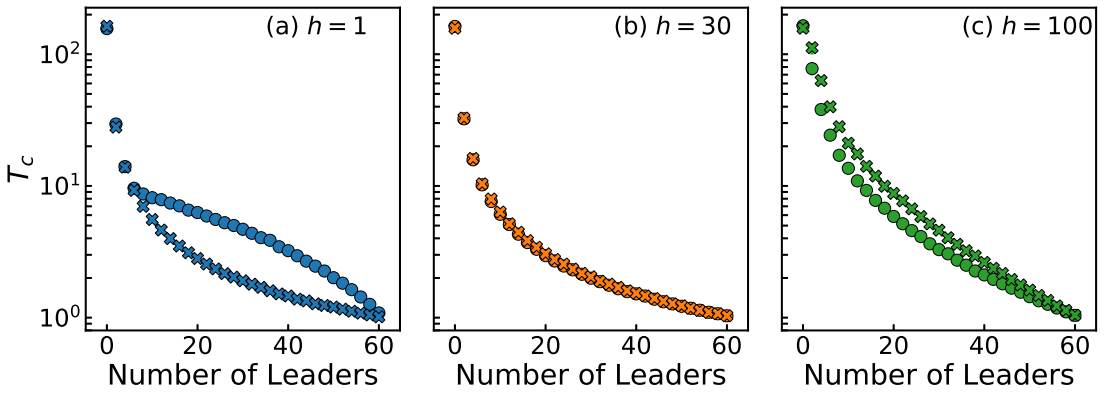

FIG. 13. Cluster correlation time for compact and extended clusters. We compare the cluster correlation time T_c between $N = 61$ trains and $Q = 4$ oligomers. The follower noise level is 36°. Circles represent the trains and Xs represent the oligomers.

significantly as a function of the position. This is reflected in the very similar velocities between the geometries at all leader numbers in the $h = 30$ gradient. In the wide $h = 100$ gradient, the cells that can most accurately sense the gradient are near the back of each cluster. Therefore, the train migrates more quickly than the packed cluster, as the more extended geometry gives access to more accurate leaders. However, the differences between the geometries are not as striking as in the sharp gradient case because the $h = 100$ uncertainty curve does not vary as dramatically with the position.

One feature of the curves considered in Fig. 11 is that the nonmonotonic behavior of the velocity as a function of the leader number is present in both geometries. We look to see if the nonmonotonicity in the $N_\ell^{\max}$ as a function of the gradient width h also is robust to geometry. Figure 12 shows that for the $Q = 4$ layered oligomer, $N_\ell^{\max}$ first increases, then decreases as a function of the gradient width h. The hexagonally packed cluster starts out with a higher $N_\ell^{\max}$ at low h because more of its cells are near the transition region. It also experiences a steeper initial decrease in $N_\ell^{\max}$ as h increases, as the uncertainty curve changes from having a minimum near the transition regime to having a minimum far from the transition regime. This change more sharply affects the oligomer cluster because it does not extend as far in the x direction as the train. Finally, each has a similar number of leaders to maximize migration in the wide gradient limit. In this regime, the follower contribution becomes important relative to the leader signal. This is reflected in the breakdown of the independent follower approximation and the similar simulated values of $N_\ell^{\max}$ for both geometries.

We also consider the cluster correlation time T_c for each geometry. The main qualitative trend that the correlation time can change dramatically from a small leader fraction to the all-leader case, is robust to the geometry, as in Fig. 13.

APPENDIX G: INTRACELLULAR NOISE

In the text we assume that the uncertainty in the leaders' direction is due entirely to effects of the stochastic binding between ligands and receptors. However, additional noise may be introduced as the cell processes the information from the bound and unbound receptors. We vary the level of intracellular noise Δ_{int} that the leaders experience. We assume that this noise is independent of the ligand-receptor noise and add it in quadrature with the directional uncertainty so that a leader's uncertainty is $\Delta_\ell = \sqrt{\Delta_\phi^2 + \Delta_{\text{int}}^2}$. As the level of intrinsic noise increases, the effect of the gradient on leadership strategy washes out, as shown in Fig. 14. However, the effects do not completely wash out until the level of intracellular noise approaches extremely high levels (∼72°).

APPENDIX H: TABLE OF PARAMETERS

The following table presents the parameters which are the same for all of the simulations. Varying the physical parameters $C_{\max}$ and N_r changes the results quantitatively but not qualitatively, as long as $C_{\max}$ is of the order of K_d and N_r is within typical values.

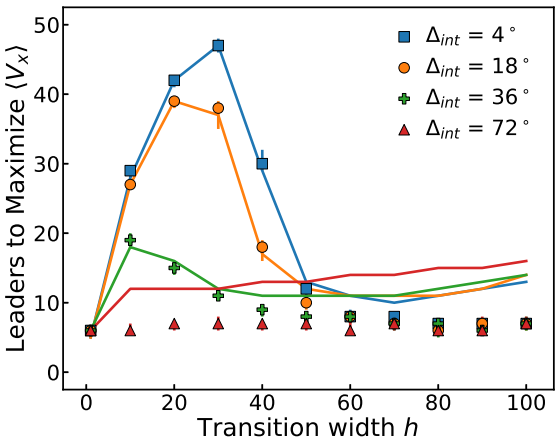

FIG. 14. Qualitative features of leadership are robust to realistic levels of intracellular noise. Increasing the intracellular noise Δ_{int} eventually washes out the effects of the gradient on the number of leaders to maximize $\langle V_x \rangle$, but for physically realistic values the qualitative features remain. Symbols represent data for $N = 50$ trains with a follower noise of $\Delta_f = 36°$; lines, the independent follower approximation predictions.

Parameter	Description	Value	Justification
$C_{\max}$	Concentration far in front of the cell cluster. Half this value is the ambient concentration of a cell at $x = 0$.	$2K_d$	Chosen so that concentrations are measurable for the cells in the cluster
N_r	Number of receptors on the surface of the cell	70 000	Cells typically have between 10 000 and 100 000 receptors [49–51].
Δt	Simulation time step	0.01	Convergence of $\langle V_x \rangle$ and $\varphi(t)$

[1] G. Malet-Engra, Y. Weimiao, A. Oldani, J. Rey-Barroso, N. S. Gov, G. Scita, and L. Dupre, Report collective cell motility promotes chemotactic prowess and resistance to chemorepulsion, Curr. Biol. **25**, 242 (2015).

[2] R. Mayor and S. Etienne-Manneville, The front and rear of collective cell migration, Nat. Rev. Mol. Cell Biol. **17**, 97 (2016).

[3] B. A. Camley, Collective gradient sensing and chemotaxis: Modeling and recent developments, J. Phys.: Condens. Matter **30**, 223001 (2018).

[4] A. Gopinathan and N. S. Gov, Cell cluster migration: Connecting experiments with physical models, Semin. Cell Dev. Biol. **93**, 77 (2019).

[5] J. Varennes and A. Mugler, Sense and sensitivity: Physical limits to multicellular sensing, migration, and drug response, Mol. Pharm. **13**, 2224 (2016).

[6] R. Mclennan, L. Dyson, K. W. Prather, J. A. Morrison, R. E. Baker, P. K. Maini, and P. M. Kulesa, Multiscale mechanisms of cell migration during development: Theory and experiment, J. Cell Sci. **125**, 2935 (2012).

[7] P. Haas and D. Gilmour, Chemokine signaling mediates self-organizing tissue migration in the zebrafish lateral line, Dev. Cell **10**, 673 (2006).

[8] T. Colak-Champollion, L. Lan, A. R. Jadhav, N. Yamaguchi, G. Venkiteswaran, H. Patel, M. Cammer, M. Meier-Schellersheim, and H. Knaut, Cadherin-mediated cell coupling coordinates chemokine sensing across collectively migrating cells, Curr. Biol. **29**, 2570 (2019).

[9] E. Donà, J. D. Barry, G. Vlentin, C. Quirin, A. Khmelinskii, A. Kunze, S. Durdu, L. Newton, A. Fernandez-Minan, W. Huber, M. Knop, and D. Gilmour, Directional tissue migration through a self-generated chemokine gradient, Nature **503**, 285 (2013).

[10] R. Mclennan, L. J. Schumacher, J. A. Morrison, J. M. Teddy, D. A. Ridenour, A. C. Box, C. L. Semerad, H. Li, W. Mcdowell, D. Kay, P. K. Maini, and R. E. Baker, Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front, Development **142**, 2014 (2015).

[11] M. Reffay, M. C. Parrini, O. Cochet-Escartin, B. Ladoux, A. Buguin, S. Coscoy, F. Amblard, J. Camonis, and P. Silberzan, Interplay of RhoA and mechanical forces in collective cell migration driven by leader cells, Nat. Cell Biol. **16**, 217 (2014).

[12] E. Theveneau and C. Linker, Leaders in collective migration: Are front cells really endowed with a particular set of skills? F1000Research **6**, 1899 (2017).

[13] C. Dambly-Chaudière, N. Cubedo, and A. Ghysen, Control of cell migration in the development of the posterior lateral line: Antagonistic interactions between the chemokine receptors CXCR4 and CXCR7/RDC1, BMC Dev. Biol. **7**, 1 (2007).

[14] S. J. Streichan, G. Valentin, D. Gilmour, and L. Hufnagel, Collective cell migration guided by dynamically maintained gradients, Phys. Biol. **8**, 045004 (2011).

[15] C. Scherber, A. J. Aranyosi, B. Kulemann, S. P. Thayer, M. Toner, O. Iliopoulos, and D. Irimia, Epithelial cell guidance by self-generated EGF gradients, Integrat. Biol.: Quant. Biosci. Nano Macro **4**, 259 (2012).

[16] G. Venkiteswaran, S. W. Lewellis, J. Wang, E. Reynolds, C. Nicholson, and H. Knaut, X generation and dynamics of an endogenous, self-generated signaling gradient across a migrating tissue, Cell **155**, 674 (2013).

[17] A. J. Muinonen-Martin, O. Susanto, Q. Zhang, E. Smethurst, W. J. Faller, D. M. Veltman, G. Kalna, C. Lindsay, D. C. Bennett, O. J. Sansom, R. Herd, R. Jones, L. M. Machesky, M. J. O. Wakelam, D. A. Knecht, and R. H. Insall, Melanoma cells break down LPA to establish local gradients that drive chemotactic dispersal, PLoS Biol. **12**, e1001966 (2014).

[18] L. Tweedy, D. A. Knecht, G. M. Mackay, and R. H. Insall, Self-generated chemoattractant gradients: Attractant depletion extends the range and robustness of chemotaxis, PLoS Biol. **14**, e1002404 (2016).

[19] D. Dalle Nogare and A. B. Chitnis, A framework for understanding morphogenesis and migration of the zebrafish posterior lateral line primordium, Mech. Dev. **148**, 69 (2017).

[20] O. Susanto, Y. W. H. Koh, N. Morrice, S. Tumanov, P. A. Thomason, M. Nielson, L. Tweedy, A. J. Muinonen-Martin, J. J. Kamphorst, G. M. Mackay, and R. H. Insall, LPP3 mediates self-generation of chemotactic LPA gradients by melanoma cells, J. Cell Sci. **130**, 3455 (2017).

[21] K. Wang, W.-J. Rappel, R. Kerr, and H. Levine, Quantifying noise levels of intercellular signals, Phys. Rev. E **75**, 061905 (2007).

[22] B. Hu, W. Chen, W.-J. Rappel, and H. Levine, Physical Limits on Cellular Sensing of Spatial Gradients, Phys. Rev. Lett. **105**, 048104 (2010).

[23] B. Hu, W. Chen, W.-J. Rappel, and H. Levine, How geometry and internal bias affect the accuracy of eukaryotic gradient sensing, Phys. Rev. E **83**, 021917 (2011).

[24] H. Levine and W.-J. Rappel, The physics of eukaryotic chemotaxis, Phys. Today **66**(2), 24 (2013).

[25] M. Ueda and T. Shibata, Stochastic signal processing and transduction in chemotactic response of eukaryotic cells, Biophys. J. **93**, 11 (2007).

[26] R. G. Endres and N. S. Wingreen, Accuracy of direct gradient sensing by single cells, Proc. Natl. Acad. Sci. USA **105**, 15749 (2008).

[27] D. Fuller, A. Groisman, W.-J. Rappel, W. F. Loomis, H. Levine, M. Adler, and W. Chen, External and internal constraints on

eukaryotic chemotaxis, Proc. Natl. Acad. Sci. USA **107**, 9656 (2010).

[28] I. Segota, S. Mong, E. Neidich, A. Rachakonda, C. J. Lussenhop, and C. Franck, High fidelity information processing in folic acid chemotaxis of Dictyostelium amoebae, J. R. Soc. Interface **10**, 20130606 (2013).

[29] G. Amselem, M. Theves, A. Bae, C. Beta, and E. Bodenschatz, Control Parameter Description of Eukaryotic Chemotaxis, Phys. Rev. Lett. **109**, 108103 (2012).

[30] B. A. Camley, J. Zimmermann, H. Levine, and W.-J. Rappel, Emergent Collective Chemotaxis without Single-Cell Gradient Sensing, Phys. Rev. Lett. **116**, 098101 (2016).

[31] B. A. Camley and W.-J. Rappel, Physical models of collective cell motility: From cell to tissue, J. Phys. D **50**, 113002 (2017).

[32] B. Szabó, G. J. Szöllösi, B. Gönci, Z. Jurányi, D. Selmeczi, and T. Vicsek, Phase transition in the collective migration of tissue cells: Experiment and model, Phys. Rev. E **74**, 061908 (2006).

[33] B. A. Camley and W.-J. Rappel, Velocity alignment leads to high persistence in confined cells, Phys. Rev. E **89**, 062705 (2014).

[34] A. Szabó, R. Ünnep, E. Méhes, W. O. Twal, W. S. Argraves, Y. Cao, and A. Czirók, Collective cell motion in endothelial monolayers, Phys. Biol. **7**, 046007 (2010).

[35] M. Basan, J. Elgeti, E. Hannezo, W.-J. Rappel, and H. Levine, Alignment of cellular motility forces with tissue flow as a mechanism for efficient wound healing, Proc. Natl. Acad. Sci. USA **110**, 2452 (2013).

[36] We study the mean velocity in the gradient direction rather than the chemotactic index, as the mean velocity is most relevant for self-generated gradients enhancing a cell cluster's motion along a fixed path [10].

[37] R. McLennan, L. J. Schumacher, J. A. Morrison, J. M. Teddy, D. A. Ridenour, A. C. Box, C. L. Semerad, H. Li, W. McDowell, D. Kay, P. K. Maini, R. E. Baker, and P. M. Kulesa, VEGF signals induce trailblazer cell identity that drives neural crest migration, Dev. Biol. **407**, 12 (2015).

[38] K. J. Cheung and A. J. Ewald, A collective route to metastasis: Seeding by tumor cell clusters, Science **352**, 167 (2016).

[39] M. L. Lalli and A. R. Asthagiri, Collective migration exhibits greater sensitivity but slower dynamics of alignment to applied electric fields, Cell. Mol. Bioeng. **8**, 247 (2015).

[40] F. J. Segerer, F. Thüroff, A. Piera Alberola, E. Frey, and J. O. Rädler, Emergence and Persistence of Collective Cell Migration on Small Circular Micropatterns, Phys. Rev. Lett. **114**, 228102 (2015).

[41] L. Coburn, L. Cerone, C. Torney, I. D. Couzin, and Z. Neufeld, Tactile interactions lead to coherent motion and enhanced chemotaxis of migrating cells, Phys. Biol. **10**, 046002 (2013).

[42] D. Cai, W. Dai, M. Prasad, J. Luo, N. S. Gov, and D. J. Montell, Modeling and analysis of collective cell migration in an in vivo three-dimensional environment, Proc. Natl. Acad. Sci. USA **113**, E2134 (2016).

[43] D. Ellison, A. Mugler, M. D. Brennan, S. H. Lee, R. J. Huebner, E. R. Shamir, L. A. Woo, J. Kim, P. Amar, I. Nemenman, A. J. Ewald, and A. Levchenko, Cell-cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis, Proc. Natl. Acad. Sci. USA **113**, E679 (2016).

[44] D. Dalle Nogare, M. Nikaido, K. Somers, J. Head, T. Piotrowski, and A. B. Chitnis, In toto imaging of the migrating Zebrafish lateral line primordium at single cell resolution, Dev. Biol. **422**, 14 (2017).

[45] A. J. Kabla, Collective cell migration: Leadership, invasion and segregation, J. R. Soc. Interface **9**, 3268 (2012).

[46] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. (Elsevier, Amsterdam, 2007).

[47] https://github.com/bcamley/leaders-hopkins

[48] B. Efron, Bootstrap methods: Another look at the jackknife, Ann. Stat. **7**, 1 (1979).

[49] R. J. W. de Wit and P. J. M. van Haastert, Binding of folates to Dictyostelium discoideum cells. Demonstration of five classes of binding sites and their interconversion, BBA Biomembranes **814**, 199 (1985).

[50] S. D. Tennenberg, F. P. Zemlan, and J. S. Solomkin, Characterization of N-formyl-methionyl-leucyl-phenylalanine receptors on human neutrophils. Effects of isolation and temperature on receptor expression and functional activity, J. Immunol. **141**, 3937 (1988).

[51] B. Wurster and U. Butz, Reversible binding of the chemoattractant folic acid to cells of Dictyostelium discoideum, Eur. J. Biochem. **109**, 613 (1980).