

Collective gradient sensing with limited positional information

Emiliano Perez Ipiña*

Department of Physics & Astronomy, Johns Hopkins University, Baltimore, MD.

Brian A. Camley†

Department of Physics & Astronomy and Biophysics, Johns Hopkins University, Baltimore, MD.

Eukaryotic cells sense chemical gradients to decide where and when to move. Clusters of cells can sense gradients more accurately than individual cells by integrating measurements of the concentration made across the cluster. Is this gradient sensing accuracy impeded when cells have limited knowledge of their position within the cluster, i.e. limited positional information? We apply maximum likelihood estimation to study gradient sensing accuracy of a cluster of cells with finite positional information. If cells must estimate their location within the cluster, this lowers the accuracy of collective gradient sensing. We compare our results with a tug-of-war model where cells respond to the gradient by polarizing away from their neighbors without relying on their positional information. As the cell positional uncertainty increases, there is a trade-off where the tug-of-war model responds more accurately to the chemical gradient. However, for sufficiently large cell clusters or sufficiently shallow chemical gradients, the tug-of-war model will always be suboptimal to one that integrates information from all cells, even if positional uncertainty is high.

I. INTRODUCTION

Eukaryotic cells may directionally migrate by sensing external cues of many types – mechanical, electrical, topographical, or chemical [1], and in these processes their accuracy is often limited by basic physical principles [2]. The best-known of these processes is chemotaxis, where cells sense and follow gradients in concentrations of chemical cues – an example of gradient sensing. Gradient sensing and directed migration are essential to many fundamental biological processes, such as immune response [3], embryonic development [4, 5], and cancer metastasis [6, 7]. In these processes, cells may migrate individually or collectively – in sheets, streams, or small clusters [8–10]. Cell cluster migration may be particularly important in cancer, where small clusters of tumor cells are associated with more harmful metastasis [11]. There are several advantages for cells to migrate as groups [12], one of these being gradient sensing. In several biological systems [5, 13, 14], groups of cells can chemotax where single cells fail to do so – cells work collectively to improve their ability to sense a gradient [15].

How this enhanced sensing arises is not fully understood, and may not have a universal mechanism [15], but two types of cell clusters seem to primarily sense using measurements of the concentration at the cluster edge. In clusters of neural crest cells responding to gradients of Sdf1, edge cells are polarized out from the cluster center by contact inhibition of locomotion [5], suggesting a model of chemotaxis as a tug-of-war by the perimeter cells [16]. A similar edge-driven mechanism is observed in gradient-sensing lymphocyte clusters [13, 17]. These observations suggest that in these cell types only

edge cells are involved in cluster chemotaxis. Why use only the edge cells to sense the gradient instead of all of them? Our initial intuition is that using only the edge cells wastes information [18]. In earlier work, we found a fundamental limit on gradient sensing accuracy [19], observing that collective chemotaxis was likely limited by cell-to-cell variability. In that model, measurements from all cells and their relative positions are used to compute the best estimate of the gradient direction.

Why would clusters use sensing mechanisms driven by edge cells, like the tug-of-war mechanism, instead of the optimal use of all cells? One hypothesis is that the benefit of the extra information is small. Another hypothesis is that there are relevant sources of noise that are not included in Ref. [19]. A major assumption implicit in our earlier work [19] is that cells “know” their position within a cluster – the best estimator of the chemical gradient involves a sum weighted by a cell’s position relative to the cluster center. Cells can measure their position within an embryo or aggregate by measuring diffusible [20] or mechanical [21] signals but this positional information is limited [20, 22–24]. Cells in *Drosophila* embryos can measure their position to the order of a cell length [20], using temporal and possibly spatial averaging of graded signals [20, 25, 26]. In this paper, we explore the optimal strategies for collective sensing of gradients when cells have limited information on their position in the cluster, showing that positional uncertainty reduces gradient sensing accuracy. We also compare our model to an extension of the tug-of-war model of [16], where only edge cells respond to chemoattractant and no positional information is required. We find that these tug-of-war strategies are optimal for small clusters or strong gradients. However, for a sufficiently large cluster the benefits of using all cells’ data will eventually outweigh the error from finite positional information.

* emperipi@jhu.edu

† bcamley@jhu.edu

II. MODELS AND RESULTS

A. Maximum likelihood estimates of gradient with limited positional information

We consider N circular cells of radius R_{cell} arranged in 2D clusters under a linear chemoattractant gradient $\mathbf{g} = g(\cos(\phi), \sin(\phi))$, see Fig. 1a. g and ϕ are the gradient magnitude or steepness and direction respectively. The chemoattractant concentration at cell i is $c(\mathbf{r}_i) = c_0(1 + \mathbf{g} \cdot (\mathbf{r}_i - \mathbf{r}_{\text{cm}}))$, where \mathbf{r}_i is the cell's position, and \mathbf{r}_{cm} is the position of the cluster center of mass, computed as $\mathbf{r}_{\text{cm}} = \frac{1}{N} \sum \mathbf{r}_i$. c_0 is the chemoattractant concentration at the cluster center of mass. Cells perform a collective measurement of the gradient by integrating individual cells' measurements of local chemoattractant concentration, which we call the “measured signal” M_i . Cells sense the concentration by binding the chemoattractant molecules to receptors at their surface, which is subject to fluctuations due to the molecules' diffusion in reaching receptors, and intrinsic ligand-receptor kinetics [26–29]. The reading concentration error of a cell at position \mathbf{r}_i with n_r receptors that bind ligand molecules with dissociation constant K_D , can be expressed as $\delta c^2(\mathbf{r}_i) = \frac{c(\mathbf{r}_i)}{n_r K_D} (c(\mathbf{r}_i) + K_D)^2$ [19, 26]. This arises solely from the stochasticity of ligand-receptor binding, propagated forward to an estimate of the concentration (Appendix A). In addition, even genetically identical cells do not respond in the same way to the chemoattractant due to cell-cell variability (CCV) resulting from fluctuations in their internal molecular machinery [19]. Then, the signal measured by cell i is

$$M_i = \frac{c(\mathbf{r}_i) + \delta c(\mathbf{r}_i)\eta_i}{\bar{c}} + \sigma_{\Delta}\xi_i, \quad (1)$$

where η_i and ξ_i are independent Gaussian noises with mean value 0 and variance 1, \bar{c} is the mean concentration over the cluster, and σ_{Δ} is the CCV standard deviation. Note that $\bar{c} = \frac{1}{N} \sum_i^N c(\mathbf{r}_i) = c_0$. For typical values of $\sigma_{\Delta} \geq 0.05$ (i.e. 5% cell-to-cell variability), ligand-receptor noise is less critical than CCV noise [19].

We want to know, given the noise in the cell cluster's measurements, what is the best possible measurement it can make of the gradient \mathbf{g} . One approach to find this gradient sensing error is to apply the maximum likelihood estimation (MLE) method [30–35]. This calculation, which we will generalize below, determines an optimal approach for a group of cells to estimate \mathbf{g} given the measured signal of Eq. (1) [19]. However, this approach implicitly assumes that the positions of the cells are known, as each measurement M_i correspond to a given position \mathbf{r}_i within the gradient.

Here we extend the MLE approach of [19] by assuming that cells have limited positional information. As a result, cells need to estimate their positions within the cluster in addition to the local concentration. We then define a “measured position” for cell i , which we call \mathbf{r}_i^* – this is the location that the cluster believes the cell i to

have. Because cells can gain positional information from a variety of complicated mechanical and chemical processes, we simplify and start with a generic assumption – that the error in measured position is normally distributed, and that cells get their positional information from a process different from the sensing concentration such that \mathbf{r}_i^* and M_i are independent. Thus, the measured position for the cell i is,

$$\mathbf{r}_i^* = \mathbf{r}_i + \mathcal{A}_i \boldsymbol{\xi}_{\mathbf{r}_i}, \quad (2)$$

where \mathcal{A}_i is a matrix such that $\boldsymbol{\Sigma}_i = \mathcal{A}_i \mathcal{A}_i^T$ is the covariance matrix of the positional errors, and $\boldsymbol{\xi}_{\mathbf{r}_i}$ is a two dimensional vector of uncorrelated normal distributed numbers. X^T denotes the transpose of X . The covariance matrix $\boldsymbol{\Sigma}_i$ sets the error in the position measurement process, analogous to the term σ_{Δ}^2 in sensing the concentration. We assume for now that $\boldsymbol{\Sigma}_i$ could be different from cell to cell.

We apply the maximum likelihood estimation (MLE) method to determine, given the measured cell positions \mathbf{r}_i^* and signal values M_i , what the cluster's best estimate of the gradient \mathbf{g} is, and how precise this measurement can be given the inevitable stochastic fluctuations in these measurements. We start by writing the probability distribution for a cell i to measure a signal M_i and a position \mathbf{r}_i^* given the gradient \mathbf{g} and its true position \mathbf{r}_i , $p(M_i, \mathbf{r}_i^* | \mathbf{g}, \mathbf{r}_i)$. As M_i and \mathbf{r}_i^* are independent, this probability function can be broken into the product of probabilities $p(M_i, \mathbf{r}_i^* | \mathbf{g}, \mathbf{r}_i) = p(M_i | \mathbf{g}, \mathbf{r}_i)p(\mathbf{r}_i^* | \mathbf{r}_i)$. Given that Eq. (1) takes the sum of two uncorrelated Gaussian random variables, which is also Gaussian, then, the probability distribution for M_i is

$$p(M_i | \mathbf{g}, \mathbf{r}_i) = \frac{1}{\sqrt{2\pi h_i}} \exp \left[-\frac{(M_i - \mu_i)^2}{2h_i} \right], \quad (3)$$

where $\mu_i = 1 + \mathbf{g} \cdot \mathbf{r}_i$, and $h_i = (\delta c_i/c_0)^2 + \sigma_{\Delta}^2$. Next, the probability distribution that a cell i measures a position \mathbf{r}_i^* given the true position \mathbf{r}_i is,

$$p(\mathbf{r}_i^* | \mathbf{r}_i) = \frac{1}{\sqrt{(2\pi)^2 |\boldsymbol{\Sigma}_i|}} \exp \left[-\frac{(\mathbf{r}_i^* - \mathbf{r}_i) \boldsymbol{\Sigma}_i^{-1} (\mathbf{r}_i^* - \mathbf{r}_i)^T}{2} \right]. \quad (4)$$

The likelihood of parameters $\mathbf{g}, \{\mathbf{r}_i\}$ given that the clusters measures values $\{M_i\}$ and $\{\mathbf{r}_i^*\}$ is $\mathcal{L}(\mathbf{g}, \{\mathbf{r}_i\} | \{M_i, \mathbf{r}_i^*\}) \equiv p(\{M_i, \mathbf{r}_i^*\} | \mathbf{g}, \{\mathbf{r}_i\})$. Assuming the measurements cell performs are independent, this likelihood then factorizes into a product over cells i , $\mathcal{L}(\mathbf{g}, \{\mathbf{r}_i\} | \{M_i, \mathbf{r}_i^*\}) = \prod_i^N p(M_i | \mathbf{g}, \mathbf{r}_i)p(\mathbf{r}_i^* | \mathbf{r}_i)$. Estimators of the gradient magnitude g and orientation ϕ can be obtained by maximizing this likelihood. However, we are more interested in the best possible accuracy for an unbiased estimator \hat{g} , which is given by the Cramér-Rao bound [30],

$$\sigma_g^2 \equiv \langle (g - \hat{g})^2 \rangle = (\mathcal{I}^{-1})_{g,g}, \quad (5)$$

where \mathcal{I}^{-1} is the inverse of the Fisher information matrix, given by $\mathcal{I}_{\alpha,\beta} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha \partial \beta} \right\rangle$, with α and β are parameters of the likelihood function, *i.e.* \mathbf{g} and $\{\mathbf{r}_i\}$.

It is possible to analytically compute the Fisher information \mathcal{I} and thus the best possible accuracy for measuring the chemical gradient in the limit of positional information in terms of a sum over cell positions and the positional information at each location. These results are presented in full detail in Appendix B. However, the results are much more intuitively understandable with a few key assumptions.

B. Simplest case: shallow gradients and constant positional error

Cells will likely have different positional uncertainties, depending on the underlying process through which cells get their positional information, such as an external or a self-generated signal. To gain understanding, we will first look at the simplest case by assuming all the cells have the same positional covariance matrix, $\Sigma_i = \Sigma$, which we also assume is isotropic $\Sigma_{ls} = \Delta r^2 \delta_{ls}$, where l, s are matrix element indexes. (We study clusters with varying positional uncertainty in Section II C). Within this section, the limited positional information is only characterized by a single number, Δr – the error in measuring a cell’s position. In addition to considering a constant isotropic error, we simplify our results by assuming that the gradient is relatively shallow – *i.e.* that each cell has the same reading concentration errors, $\delta c_i^2 \approx \overline{\delta c}^2 = \frac{c_0(c_0+K_D)^2}{n_r K_D}$. Finally, we consider that our cluster of cells is roughly circular: the isotropy of the problem then makes $\sum_i \delta r_{x_i}^2 \approx \sum_i \delta r_{y_i}^2 \approx \frac{1}{2} \sum_i |\delta \mathbf{r}_i|^2$, and that $\sum_i \delta r_{x_i} \delta r_{y_i} \approx 0$. (In practice, for the calculations we present here, we show results for hexagonally packed clusters of layers of cells, as in [16, 19]). With all this, we can simplify the results of Appendix B to find the best possible error in measuring the gradient magnitude g as:

$$\sigma_g^2 \approx \frac{1}{\chi} \left(\bar{h} + (g\Delta r)^2 \right), \quad (6)$$

where $\bar{h} = \overline{\delta c}^2 / c_0^2 + \sigma_\Delta^2$ and $\chi = \frac{1}{2} \sum_i |\delta \mathbf{r}_i|^2$. χ here is a geometric factor that depends on the shape and size of the cluster – essentially like a moment of inertia about the cluster’s centroid. We expect χ to scale as $\chi \sim R_{\text{cluster}}^4$ [19] for roughly circular clusters. We find this by approximating the sum over cells as an integral over the surface of a circular cluster of radius R_{cluster} , $\sum_i R_{\text{cell}}^2 \approx \int_{R_{\text{cluster}}} d^2 r$, so $\chi = \frac{1}{2} \sum_i |\delta \mathbf{r}_i|^2 \sim R_{\text{cell}}^{-2} \int_{R_{\text{cluster}}} d^2 r r^2 \sim R_{\text{cluster}}^4$.

The uncertainty in the gradient magnitude given in Eq. (6) also controls the uncertainty of estimating the gradient direction: the gradient direction error is equal

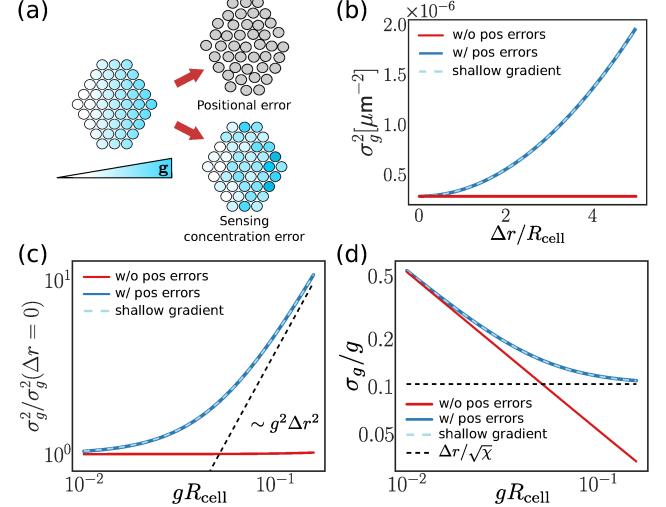


FIG. 1. (a) Cluster under a linear gradient and a representation of the two sources of uncertainties: sensing concentration and positional errors. (b) Gradient sensing error increases with positional uncertainty. (c)–(d) **Positional uncertainties becomes more dominant in steeper gradients.** (c) Gradient sensing error normalized by its value in the absence of positional errors, $\sigma_g^2(\Delta r = 0)$, as a function of the gradient steepness (blue solid line). The black dashed-line indicates the limit where the gradient sensing error is dominated by positional errors and it follows $\propto g^2 \Delta r^2$. (d) Gradient sensing relative error σ_g/g as a function of the gradient steepness. Parameters: number of cells $N = 37$ (corresponding to an hexagonal cluster of $Q = 3$ layers), $R_{\text{cell}} = 10 \mu\text{m}$, (b) $g = 0.005 \mu\text{m}^{-1}$, (c)–(d) $\Delta r = 10 \mu\text{m}$.

to the gradient magnitude relative error,

$$\sigma_\phi = \frac{\sigma_g}{g}. \quad (7)$$

See Appendix B for detailed derivation of Eqs. (6)–(7).

How does the presence of positional uncertainties affect gradient sensing? In Fig. 1(b) we show the gradient sensing error, σ_g^2 , as a function of the positional uncertainty Δr , using both the full computation of Eq. (5) (see Appendix B) and the shallow gradient approximation from Eq. (6). Unsurprisingly, positional uncertainties increase gradient sensing error – the more uninformed cells are about their positions, the worse an estimate the cluster makes. This added error from positional uncertainty increases as Δr^2 in the shallow gradient limit (Eq. (6)). We see in Fig. 1 that the additional uncertainty from finite positional information is significant when Δr is of order of the cell size.

Eq. (6) also tells us there are two terms controlling the estimate error. The first, \bar{h} , is the contribution from the cells reading and reporting their local concentration. The second term, is the positional uncertainty contribution to the gradient sensing error, controlled by $g\Delta r$. The relative weight of these two terms allows to identify two regimes in which the gradient sensing is limited in one case by the reading concentration fluctuations,

$(g\Delta r)^2 \ll \bar{h}$, and on the other case by the positional uncertainties, $(g\Delta r)^2 \gg \bar{h}$. When positional uncertainties are smaller than reading concentration errors, we recover the results from [19], $\sigma_g^2 \approx \frac{1}{\chi} \bar{h}$. In the other extreme, $(g\Delta r)^2 \gg \bar{h}$, gradient sensing is primarily limited by positional information, $\sigma_g^2 \approx \frac{1}{\chi} g^2 \Delta r^2$. Interestingly, all of the information about the cluster's shape and size is in the multiplicative factor χ . While larger clusters are significantly more accurate at sensing the gradient, as the factor $\chi = \frac{1}{2} \sum_i^N |\delta \mathbf{r}_i|^2$ is strongly dependent on cluster size and cell number, larger clusters are not more strongly affected by the presence of positional error.

The gradient steepness g , controls the relative importance of positional information: the steeper the gradient is, the larger the contribution from the positional uncertainty. In Fig. 1(c) we show the gradient sensing error σ_g^2 normalized by the bound in absence of positional uncertainty $\sigma_g^2(\Delta r = 0)$ increases with the gradient steepness g . In steeper gradients, small errors in measuring a cell's position lead to larger differences between the concentration at the true position, $c(\mathbf{r}_i)$, and at the estimated position, $c(\mathbf{r}_i^*)$. On the other hand, this also implies, that in the limit of $g \rightarrow 0$, where error bounds are most important (where single cells fail to follow a gradient while clusters do), that positional information does not play a significant role.

In analogy to single-cell gradient sensing, it would be natural to assume that the relative error, σ_g/g , which controls the angular uncertainty σ_ϕ , always gets better with the increasing g – i.e. that cells can better measure the angle to the gradient direction if the gradient is steeper. This is true for clusters without positional error [19]: σ_g/g is plotted as the red line in Fig. 1(d). However, in the presence of positional uncertainties, the relative error is limited at large g , converging to a constant value $\Delta r/\sqrt{\chi}$ in the case of a shallow gradient (Fig. 1(d)). This surprise occurs because the increasing error due to positional uncertainty $\sigma_g^2 \sim g^2$ in Eq. (6) exactly balances the g^2 in the denominator of σ_g^2/g^2 .

C. Non-uniform positional information

In the above section we assumed that all cells have the same positional errors. However, positional uncertainty is not necessarily uniform within the cluster. If cells get their positional information from a secreted factor or mechanical signal, the distance from the source may influence the accuracy Δr (see, e.g. [36]). In addition, obtaining positional information may have an associated cost and therefore, it may be optimal for the cell cluster to only have some cells measure their location. We study a prototypical example of both cell specialization and positional errors varying from cell to cell by introducing a second type of cells which have more positional information. These “informed” cells have a positional error Δr_{inf} that is smaller than “normal” cells, $\Delta r_{\text{inf}} \leq \Delta r$. Then,

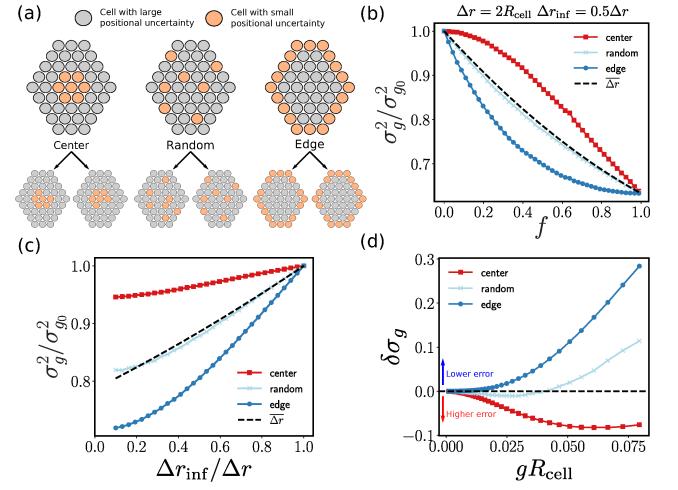


FIG. 2. Gradient sensing errors for a cluster with a fraction f of informed cells. (a) Informed cells are distributed over the cluster in three distinct manners: closer to the cluster's center, randomly, or closer to the cluster's edge. Underneath each example of cluster distribution we show two of the multiple equivalent ways of distributing the informed cells. (b)–(c) Gradient sensing variance, σ_g^2 , normalized by the case of non informed cells, $\sigma_{g0}^2 = \sigma_g^2(f=0)$ as a function of the fraction of informed cells (b) and the ratio of the two types of cell positional uncertainties (c). (d) Relative difference between informed cluster and the equivalent constant-positional-error cluster, $\delta\sigma_g = \frac{\sigma_g(\bar{\Delta}r) - \sigma_g(\Delta r_{\text{inf}}, \Delta r)}{\sigma_g(\Delta r_{\text{inf}}, \Delta r)}$, as a function of the gradient steepness. $\delta\sigma_g > 0$ ($\delta\sigma_g < 0$) indicates a lower (higher) gradient sensing error for the cluster with the informed cells compared to the uniform positional error cluster. The dashed black line indicates $\delta\sigma_g = 0$. For (b)–(c)–(d) color codes are the same. Solid lines with symbols represent the three different distribution patterns: Center (red squares), Random (light blue crosses), and Edge (dark blue circles). The dashed black line represents the equivalent one type of cells cluster with positional uncertainty equal to $\bar{\Delta}r = f\Delta r_{\text{inf}} + (1-f)\Delta r$. Parameters: cluster size $N = 91$ ($Q = 5$ layers in the hexagonal cluster), $\Delta r = 2R_{\text{cell}}$, (b) $\Delta r_{\text{inf}} = 0.5\Delta r$, (c) $f = 0.25$, (d) $f = 0.25$, $\Delta r_{\text{inf}} = 0.2\Delta r$, and $\bar{\Delta}r = 20\mu\text{m}$. Curves show an average of 100 realizations.

we distribute a fraction f of informed cells at different positions over the cluster to see the effect of positional uncertainty localization. We follow three different types of distributions: informed cells are distributed (i) at the edge, (ii) at the center, or (iii) randomly over the cluster, see Fig. 2(a). We use the general solution for the gradient sensing error given by Eqs. (B4)–(B6). We note that there are often many cells that are equidistant from the cluster center or edge, so unless a layer of the hexagonal packing is completely filled, there will be multiple possible configurations at a particular fraction of informed cells f (Fig. 2a, bottom row). To address this, we sample 100 realizations for each distribution. Fig. 2 shows an average over these realizations. Gradient sensing has the lowest error when informed cells are closer to the edges and the largest when they are placed closer to the cen-

ter of mass. This is illustrated in Fig. 2(b), that shows the gradient sensing error normalized by the case $f = 0$, ($\sigma_g^2 = \sigma_g^2(f = 0)$), as a function of the fraction of informed cells. For all three types of distributions, increasing the fraction of informed cells reduces the error of g . Gradient sensing error takes the same value for the three distributions, in the extreme cases of no informed cells ($f = 0$) and all cells informed ($f = 1$), but following different paths in between. Informed cells distributed close to the edge lead to the lowest sensing errors of the three distributions. Similarly, reducing the positional uncertainty of the informed cells results in higher accuracy in gradient sensing, which is again greater when informed cells are distributed closer to the edge (Fig. 2(c)).

To compare between clusters with informed cells and cluster with a single cell type population, we introduce the “equivalent” constant-positional-error cluster, in which all the cells have the average positional error $\overline{\Delta r} = (1-f)\Delta r + f\Delta r_{\text{inf}}$ (dashed black line in Fig. 2(b)–(d)). Results show that placing informed cells closer to the edge enhances gradient sensing when compared to the equivalent cluster, while doing it closer to the cluster’s center or randomly results in a worse or similar performance, respectively. To further explore the role of informed cells and their distribution inside the cluster, in Fig. 2(d) we shows the relative difference between the gradient sensing errors for the cluster with informed cells and its constant-positional-error equivalent, $\delta\sigma_g = \frac{\sigma_g(\overline{\Delta r}) - \sigma_g(\Delta r_{\text{inf}}, \Delta r)}{\sigma_g(\Delta r_{\text{inf}}, \Delta r)}$, as a function of the gradient steepness. Results show that increasing the gradient steepness enhances the trends described before, in which the benefit of reduced positional error is larger for cells at the cluster edge. This observation agrees with our previous results that showed that positional errors have a higher effect over gradient sensing in steeper gradients. Interestingly, in steeper gradients, even randomly distributed informed cells perform better than the equivalent cluster (light blue line in Fig. 2(d)). The reason is that by randomly distributing the informed cells, a few will lie close to the edge of the cluster, and those are the ones that dominate and end up having a major weight in sensing the gradient.

Together, these results show that positional information becomes more relevant at the edge of the cluster and is even more significant in steeper gradients. Positional information of edge cells plays a major contribution to gradient sensing in that even having a few randomly located informed cells on the edge can significantly decrease gradient sensing errors. If determining cell positions is costly, our results suggest that instead of maximizing positional information uniformly throughout the cluster, cells can improve the gradient sensing accuracy by prioritizing positional information of edge cells.

We argued that informed cells are generally most beneficially placed at the cluster edge. How does gradient sensing accuracy change if we place these cells anisotropically? We study the change in the errors for gradient

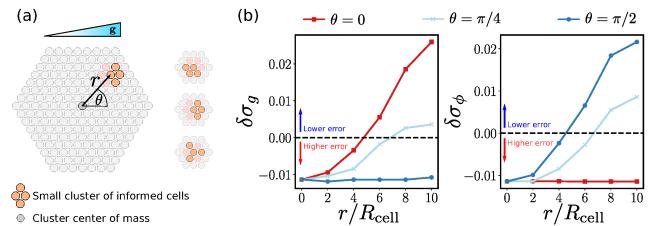


FIG. 3. (a) Schematic cell cluster with a group of four informed cells located at a distance r from the cluster’s center and direction θ with respect to the chemoattractant gradient. The small clusters on the right represent different small cluster configurations. (b) Relative gradient sensing steepness (left panel) and direction, $\delta\sigma_\phi = \frac{\sigma_\phi(\overline{\Delta r}) - \sigma_\phi(\Delta r_{\text{inf}}, \Delta r)}{\sigma_\phi(\Delta r_{\text{inf}}, \Delta r)}$, (right panel) errors when the informed group of cells is displaced in different directions: $\theta = 0$ (red line and squares), $\theta = \pi/4$ (light blue line and crosses) and $\theta = \pi/2$ (dark blue line and circles). Same as in Fig. 2, $\delta\sigma_{g,\phi} > 0$ ($\delta\sigma_{g,\phi} < 0$) represent a lower (higher) error relative to the constant-positional-error equivalent cluster. The dashed black line indicates $\delta\sigma_{g,\phi} = 0$ and also represent the equivalent cluster without informed cells and positional uncertainty $\overline{\Delta r}$. Parameters: cluster size $N = 91$ and $\Delta r_{\text{inf}} = 0.5\Delta r$. Curves show an average of 100 realizations.

magnitude g and direction ϕ when a small group of four “informed cells” are displaced from the center to the edge of the cluster in different directions, see Fig. 3(a). We note that for this case, $\sigma_\phi \neq \sigma_g/g$, due to anisotropy; we use the full analytical solution of Appendix B to compute the bounds σ_g and σ_ϕ . As in the previous case, there are multiple equivalent ways to place a cluster of informed cells at a position (r, θ) , see Fig. 3(a). To account for such variability, we averaged over 100 realization for each condition pair (r, θ) . We find that the errors for the estimators of the gradient magnitude and direction evolve differently depending on the orientation of this cluster of cells, Fig. 3(b). Moving the small informed cluster along the axis parallel to the gradient direction, $\theta = 0$, leads to an improvement in sensing the gradient steepness g . By contrast, moving along the axis perpendicular to the gradient, $\theta = \pi/2$, sensing the orientation of the gradient is improved. Finally, there is a compromise situation when moving along $\theta = \pi/4$, where both gradient steepness and direction sensing are improved, but to a lesser extent. These results support again that positional information is more significant at the edge of the cluster, but also that the orientation with respect to the gradient matters. In chemotaxis, for instance, cells need to sense the direction rather than the steepness of the gradient. Therefore, a cluster might strive to locate cells at the edge of the axis perpendicular to the gradient direction to enhance its ability to chemotax. However, this could only work in the case that the cluster has previous knowledge of the gradient direction. Our findings are similar in concept to the ideas reported in [32] for single cells, in which elongated cells sense the gradient steepness and direction with different accuracies depending on the ori-

tation of the cell. Note that in comparing Fig. 3(b) panels that $\delta\sigma_g(0) \neq \delta\sigma_\phi(\pi/2)$ and $\delta\sigma_g(\pi/4) \neq \delta\sigma_\phi(\pi/4)$. This is due to hexagonal clusters not having perfect circular symmetry, being more elongated, in our case, in the direction parallel to the gradient.

D. Collective chemotaxis without positional information: tug-of-war model

In our above approach, a group of cells senses a chemical gradient collectively by making local measurements of the chemoattractant concentration and and their positions. Then, by applying the maximum likelihood estimator method, we find the best possible measurement of the gradient the group of cells can make. However, we showed that this measurement is constrained by the limited positional information available to the cells. Here, we introduce a different mechanism by which cells sense and follow the gradient without needing cells to measure their locations, based on the model in [16], which we refer to as *tug-of-war model*. An interesting feature of this model, which makes it different from the previous one, is that it does not depend on the positions of the cells within the cluster, but only on the direction to the nearest cells. In the tug-of-war model, cells interact by contact inhibition locomotion (CIL) [37, 38], in which contacting cells polarize away from each other. This interaction dynamic leads to a cluster where only cells at the edge are polarized [5, 13, 16], see Fig. 4(a). Cells on the inside of the cluster, completely surrounded by neighboring cells, do not polarize, whereas cells on the edges, which only see neighboring cells on the interior side, polarize away from the cluster. Another important property of the model is that cell CIL is modulated by the local chemoattractant concentration. In this way, those cells at the front (up the gradient) of the cluster polarize and pull stronger than the ones on the back (down the gradient), resulting in a net movement toward up the gradient in a tug-of-war like dynamic. This mechanism is illustrated in Fig. 4a.

Our goal in studying the tug-of-war model will be to extend the results of [16] so that they can be directly compared with our maximum likelihood results above, which requires us to include fluctuations in concentration due to receptor-ligand binding as well as cell-to-cell variability. Following [16], we start from the assumption that cells behave as stochastic, self-propelled particles. The motility of cell i results from the balance between the intracellular forces, \mathbf{F}_{ij} , due to cell-cell adhesion and volume exclusion, and the cell's polarity \mathbf{p}_i ,

$$\partial_t \mathbf{r}_i = \mathbf{p}_i + \sum_{i \neq j} \mathbf{F}_{ij}. \quad (8)$$

Here, we define the polarity vector of cell i , \mathbf{p}_i , as the velocity the cell would have if no other cells or other forces were pushing on it – this reflects both the direction and the strength of its self-propulsion. Computing the

cluster centroid velocity $\mathbf{v}_c = \frac{1}{N} \sum_{i=1}^N \partial_t \mathbf{r}_i$ by summing over all cells in Eq. (8) and noting that $\mathbf{F}_{ij} = -\mathbf{F}_{ji}$, we find

$$\mathbf{v}_c = \frac{1}{N} \sum_i \mathbf{p}_i. \quad (9)$$

Cell polarity obeys the differential equation,

$$\partial_t \mathbf{p}_i = -\frac{1}{\tau_p} \mathbf{p}_i + \sigma_p \boldsymbol{\epsilon}_i(t) + \beta_i \mathbf{q}_i, \quad (10)$$

where τ_p is the characteristic time it takes for the polarity to relax to its steady state value, and σ_p is the magnitude of a fluctuating noise that can drive the polarity away from its steady-state value. $\boldsymbol{\epsilon}_i(t)$ is a vector Gaussian Langevin noise for cell i that fulfills $\langle \boldsymbol{\epsilon}_i(t) \rangle = 0$ and $\langle \boldsymbol{\epsilon}_{i\mu}(t) \boldsymbol{\epsilon}_{j\nu}(t') \rangle = \delta_{\mu\nu} \delta_{ij} \delta(t - t')$, with i, j cell indexes, while μ, ν are indexes for the Cartesian coordinates x, y .

The last term in Eq. (10) comes from the CIL interactions, where β_i is the cell's susceptibility to CIL, and the vector \mathbf{q}_i is the resulting direction of contact interaction of cell i and its neighbors, $\mathbf{q}_i = \sum_{i \sim j} (\mathbf{r}_i - \mathbf{r}_j) / |\mathbf{r}_i - \mathbf{r}_j|$, where $i \sim j$ represent the sum over the cell's neighbors, defined as those cells within a distance $2.1R_{\text{cell}}$. Note that $\mathbf{q}_i \approx 0$ for interior cells and $\mathbf{q}_i \neq 0$ for edge cells. The tug-of-war model of [16] assumed that susceptibility of the cell i is proportional to chemoattractant concentration, $\beta_i = \bar{\beta} c(\mathbf{r}_i, t)$, leading to cells with higher concentrations c becoming more strongly polarized away from their neighbors, as sketched in Fig. 4a. Here, we must handle the variability in the *measured* concentration due to both ligand-receptor binding and cell-to-cell variability. We assume that cells polarize in response to what they believe the chemoattractant concentration to be, $M_i \bar{c}$, instead of the true concentration value at the cell positions, $c(\mathbf{r}_i, t)$. This will add an additional source of noise to Eq. (10), which we will now write explicitly. If the susceptibility of cell i is $\beta_i(t) = \bar{\beta} c_0 M_i(t)$, we can write $M_i(t) = c(\mathbf{r}_i, t)/c_0 + \Xi_i(t)$. Here, $\Xi_i(t) = \delta c_i/c_0 \zeta_i^c(t) + \sigma_\Delta \zeta_i^\Delta(t)$ is the noise in M_i . Note that, unlike Eq. (1), we must specify the time-dependence of the errors due to ligand-receptor binding and cell-to-cell variation. We explicitly introduce $\zeta_i^c(t)$ to measure the fluctuations due to concentration sensing and $\zeta_i^\Delta(t)$ for the fluctuations in cell-to-cell variability. We can characterize the time scale over which the errors due to receptor-ligand concentration and cell-to-cell variability are persistent by the correlation functions $\langle \zeta_i^c(t) \zeta_i^c(0) \rangle = C_c(t)$ and $\langle \zeta_i^\Delta(t) \zeta_i^\Delta(0) \rangle = C_\Delta(t)$. Then we have $\langle \Xi_i(t) \rangle = 0$ and $\langle \Xi_i(t) \Xi_j(0) \rangle = (\delta c_i^2/c_0^2 C_c(t) + \sigma_\Delta^2 C_\Delta(t)) \delta_{ij}$. Next, we insert the expression for the polarity susceptibility $\beta_i(t) = \bar{\beta} c(\mathbf{r}_i) + \beta c_0 \Xi_i(t)$, into Eq. (10) to arrive at

$$\partial_t \mathbf{p}_i = -\frac{1}{\tau_p} (\mathbf{p}_i - \gamma c(\mathbf{r}_i, t) \mathbf{q}_i) + \sigma_p \boldsymbol{\epsilon}_i(t) + \bar{\beta} c_0 \mathbf{q}_i \Xi_i(t), \quad (11)$$

where we have defined $\gamma = \tau_p \bar{\beta}$.

We can solve Eq. (11) by directly integrating it. If we assume that τ_p is small enough for the polarity to relax

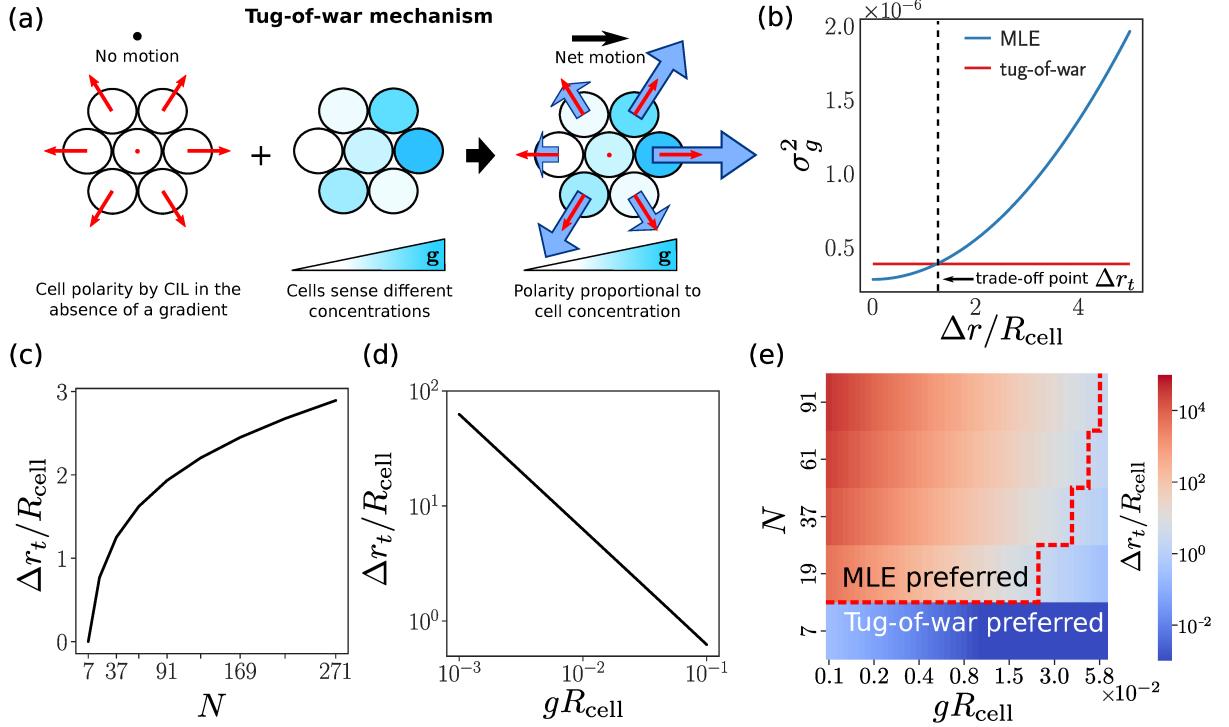


FIG. 4. **Tug-of-war model** (a) Scheme representation of the tug-of-war model. (b) Gradient sensing variance as a function of the positional uncertainty for the tug-of-war model (red solid line) and the MLE model (blue solid line). The black dashed line marks the crossover and the trade-off positional uncertainty between the two models. (c)–(d) Trade-off value as a function of: (c) the cluster number of cells, and (d) the gradient steepness. (e) Phase diagram of the trade-off values for different cluster's sizes and gradient steepness. The dashed red line indicates a boundary given by $\Delta r_t = 2R_{\text{cell}}$. Positional errors, Δr , and trade-off values, Δr_t , are re-scaled in of cell's radius. Parameters: (b) same as Fig. 1(b), (c) $g = 0.005 \mu\text{m}^{-1}$, and (d) $N = 37$ ($Q = 3$).

to its steady-state solution before the cluster reorganizes, so the contact vectors \mathbf{q}_i are fixed, then \mathbf{p}_i is given by

$$\mathbf{p}_i(t) = \boldsymbol{\mu}_i + \sigma_p \int_0^t dt' e^{-(t-t')/\tau_p} \boldsymbol{\epsilon}_i(t') + \bar{\beta} c_0 \mathbf{q}_i \int_0^t dt' e^{-(t-t')/\tau_p} \Xi_i(t') \quad (12)$$

where $\boldsymbol{\mu}_i = \gamma c(\mathbf{r}_i) \mathbf{q}_i$ and we chose initial conditions $\mathbf{p}_i(0) = \boldsymbol{\mu}_i$. From Eq. (12), it is possible to compute the mean and variance of $\mathbf{p}_i(t)$. We assume that correlations decay exponentially, so $C_s(t) = e^{-t/\tau_s}$, with τ_s the correlation time and the index $s = \{c, \Delta\}$. Our final results will not be crucially dependent on τ_c , τ_Δ , so this is primarily a convenience to make the analytical calculations simple. Then, recalling that $\boldsymbol{\epsilon}_i(t)$ is a Gaussian uncorrelated noise with mean zero, we compute the mean and the covariance of $\mathbf{p}_i(t)$ (see Appendix C),

$$\langle \mathbf{p}_i(t) \rangle = \boldsymbol{\mu}_i, \\ \langle (p_{\mu_i}(t) - \langle p_{\mu_i} \rangle)(p_{\nu_i}(0) - \langle p_{\nu_i} \rangle) \rangle = \frac{\sigma_p^2 \tau_p}{2} \delta_{\mu\nu} \left(1 - e^{-2t/\tau_p} \right) + \bar{\beta}^2 q_{\mu_i} q_{\nu_i} (\delta c_i^2 T(t, \tau_c) + c_0^2 \sigma_\Delta^2 T(t, \tau_\Delta)),$$

where

$$T(t, \tau_s) = \frac{\tau_p^2}{1 + v_s} \left(1 + \frac{1}{1 - v_s} \times \left((1 + v_s) e^{-2t/\tau_p} - 2e^{-(1+v_s)t/\tau_p} \right) \right),$$

and $v_s = \frac{\tau_p}{\tau_s}$. Note that typical values for the polarization relaxation time τ_p are ~ 20 minutes [16]. This is much longer than the typical correlation times for the chemoattractant–cell receptors binding dynamics which is on the order of the $\tau_c \sim 1$ s [19]. However, recall that cell concentration sensing is limited by CCV which has a much slower correlation time than can reach up to 48 h in human cells [19, 39]. To calculate the variance of \mathbf{p}_i at a reasonable steady-state, then, we should think of the CCV noise as constant in time, and wait for times t much longer than the other relaxation times. Explicitly, this means $\tau_\Delta \gg t \gg \tau_p \gg \tau_c$. In this limit, $T(t, \tau_\Delta) \approx \frac{\tau_p^2}{1 + \tau_p/\tau_\Delta} \approx \tau_p^2$ and $T(t, \tau_c) \approx \frac{\tau_p^2}{1 + \tau_p/\tau_c}$. As a result, the polarity covariance is $\langle (p_{\mu_i} - \langle p_{\mu_i} \rangle)(p_{\nu_i} - \langle p_{\nu_i} \rangle) \rangle = V_{\mu\nu}^i$, where $V_{\mu\nu}^i = \left(\gamma^2 c_0^2 h_{T_i} q_{\mu_i} q_{\nu_i} + \frac{\sigma_p^2 \tau_p}{2} \delta_{\mu\nu} \right)$ and $h_{T_i} = \frac{\delta c_i^2 / c_0^2}{1 + \tau_p/\tau_c} + \frac{\sigma_\Delta^2}{1 + \tau_p/\tau_\Delta} \approx \sigma_\Delta^2$. Essentially, the variance of the polarity has two terms: the time-averaged

noise from measuring concentration, which is proportional to h_{T_i} – which we have named to make clear it is effectively a time average of h_i . The second arises from the added noise σ_p in Eq. (10), which is a source of noise not included in the maximum likelihood model.

Then, with the distribution of polarities in hand, from Eq. (9) we proceed to compute the expected value and variance of the cluster velocity. In the case of the expected value, pointing out that $\sum_i^N \mathbf{q}_i = 0$, we arrive at the following simple expression,

$$\langle \mathbf{v}_c \rangle = \gamma c_0 \mathcal{M} \mathbf{g}, \quad (13)$$

where $\mathcal{M}_{\mu\nu} = \frac{1}{N} \sum_i^N \delta r_{\mu_i} q_{\nu_i}$. For the variance we have,

$$\Delta v_{c_{\mu\nu}}^2 = \frac{\gamma^2 c_0^2}{N^2} \sum_i^N h_{T_i} q_{\mu_i} q_{\nu_i} + \frac{\tau_p \sigma_p^2}{2N} \delta_{\mu\nu}. \quad (14)$$

The first term in this equation reflects the variance in the velocity arising from the fluctuations in concentration sensing h , and the second term reflects the fluctuations in polarity σ_p that affect every cell – an added noise from randomness in the motility of the cell. As the concentration is only important for edge cells where $\mathbf{q} \neq 0$, the first term scales as N_{edge}/N^2 , where N_{edge} is the number of edge cells – so we expect the first term to scale as $\sim N^{-3/2}$ for roughly circular clusters with $N_{\text{edge}} \sim \sqrt{N}$. The second term in Eq. (14) arises from averaging N independent random fluctuations, and therefore scales as N^{-1} .

Now, we ask the question, what is the gradient sensing error for this tug-of-war model? Here we are looking for an equation analogous to Eq. (6). We use Eq. (13) and Eq. (14) to propagate errors and compute the gradient sensing error for the tug-of-war model (Appendix C),

$$\sigma_g^2 \approx \frac{2}{\left(\sum_{i=1}^N \mathbf{q}_i \cdot \delta \mathbf{r}_i \right)^2} \left(\sum_{i=1}^N |\mathbf{q}_i|^2 h_{T_i} + \frac{\sigma_p^2 \tau_p}{\gamma^2 c_0^2} N \right), \quad (15)$$

under the assumption of isotropic clusters, as in Eq. (6).

How can we make a consistent comparison between the tug-of-war model and the maximum likelihood estimation? There are two key differences. First, the tug-of-war model includes an additional source of noise – the polarity noise σ_p representing fluctuations in cell polarity arising from factors outside of concentration sensing. Second, the tug-of-war model, as a dynamical model, explicitly includes an average over a characteristic time τ_p , while the maximum likelihood estimate is based on a single snapshot of the measured concentration. The first issue is simple to deal with: since the MLE method returns a lower bound for the gradient sensing error, it is most comparable to deal with the best possible situation for the tug-of-war model, *i.e.*, no polarity fluctuations, $\sigma_p = 0$. The issue of time-averaging is less clear. Eq. (15) depends on $h_{T_i} = \frac{\delta c_i^2 / c_0^2}{1 + \tau_p / \tau_c} + \frac{\sigma_\Delta^2}{1 + \tau_p / \tau_\Delta}$. To compare with the instantaneous snapshot, we would have to

choose $\tau_p \ll \tau_c, \tau_\Delta$, in which case $h_{T_i} \approx h_i$. However, this is inconsistent with our estimates above where $\tau_p \sim 20$ minutes and $\tau_c \sim 1$ second. In practice, however, this distinction is not very important, because both h_{T_i} and h_i are dominated by the CCV noise, so $h_{T_i} \approx h_i \approx \bar{h} \approx \sigma_\Delta^2$. We will then directly assume that $h_{T_i} = h_i = \bar{h}$, so that the tug-of-war and MLE results are directly comparable [40]. With these results, we can compute the uncertainty in the estimation of g arising from the tug-of-war model as:

$$\sigma_g^2 \approx \frac{1}{\chi_{\text{tow}}} \bar{h}, \quad (16)$$

where $\chi_{\text{tow}} = \frac{\left(\sum_i^N \delta \mathbf{r}_i \cdot \mathbf{q}_i \right)^2}{2 \sum_i^N |\mathbf{q}_i|^2}$. Note how similar Eq. (16) is to Eq. (6) when $\Delta r = 0$. The only difference is in the geometrical factors χ and χ_{tow} . As in Eq. (6), the factor χ_{tow} contains all the information about the cluster’s shape, size, and number of particles, while $\bar{h} = \bar{c}^2 / c_0^2 + \sigma_\Delta^2 \approx \sigma_\Delta^2$ measures the variability of sensing and responding to chemoattractant concentration. The prefactor χ_{tow} involves sums over the CIL directional vectors \mathbf{q}_i , which are only non zero on the edges cells, supporting again that only those cells contribute to sensing the gradient in the tug-of-war model.

We now want to compare the tug-of-war model and the MLE method to see which of them, and under what conditions, senses the gradient more accurately. Fig. 4(b) shows the gradient sensing error as a function of positional uncertainties for both models. In this case, since the tug-of-war model does not depend on positional uncertainties its gradient sensing error is constant. In absence of positional errors, the MLE method makes a better estimation of the gradient than the tug-of-war model, since this last one only uses information from the edge cells in contrast with the MLE model that uses all the cells. However, as cells become less informed about their positions, a trade-off occurs where, beyond a certain positional uncertainty value, Δr_t , the tug-of-war model turns out to be a better estimator of the gradient. This crossover exists since the tug-of-war model does not need to estimate cell positions and is therefore independent of positional uncertainties. Under the shallow gradient approximation, from Eq. (6) and Eq. (16), we can find analytically the crossover between both models,

$$\Delta r_t \approx \frac{\sqrt{\bar{h} \left(\frac{\chi}{\chi_{\text{tow}}} - 1 \right)}}{g}. \quad (17)$$

Using maximum likelihood estimation, increasing the cluster’s size can compensate for limitations in cells positional information, see Fig. 4(c). This result can be easily predicted from Eq. (17), by considering the scaling of the geometrical factors χ and χ_{tow} with the cluster radius, R_{cluster} . We showed in Section II B that $\chi \sim R_{\text{cluster}}^4$ for a roughly circular cluster. For the tug-of-war model, $\chi_{\text{tow}} = \frac{\left(\sum_i^N \delta \mathbf{r}_i \cdot \mathbf{q}_i \right)^2}{2 \sum_i^N |\mathbf{q}_i|^2}$. For a roughly circular cluster, \mathbf{q}_i will be along the direction $\delta \mathbf{r}_i$ for cells

at the edge, and zero for cells in the interior. Edge cells will have roughly the same magnitude of \mathbf{q}_i , $|q|$. Thus, $\delta\mathbf{r}_i \cdot \mathbf{q}_i \approx |q|R_{\text{cluster}}$ for edge cells, and $\chi_{\text{tow}} \approx (N_{\text{edge}}|q|R_{\text{cluster}})^2/2N_{\text{edge}}|q|^2 \sim N_{\text{edge}}R_{\text{cluster}}^2$. As, for a roughly circular cluster, we expect the number of cells on the edge N_{edge} to be proportional to $2\pi R_{\text{cluster}}$, we see $\chi_{\text{tow}} \sim R_{\text{cluster}}^3$. We see then that for the MLE model, the gradient sensing accuracy $\sigma_g^2 \sim R_{\text{cluster}}^{-4}$, while for the tug-of-war model, $\sigma_g^2 \sim R_{\text{cluster}}^{-3}$ – because the tug-of-war model uses only the edge cells, its accuracy grows more slowly with the cluster size. Then, $\chi/\chi_{\text{tow}} \propto R_{\text{cluster}}$ and we get from Eq. (17) that the Δr_t shifts to larger values with the cluster size, ($\Delta r_t \propto R_{\text{cluster}}^{1/2}$).

On the other hand, the gradient steepness plays an opposite role in shifting the trade-off values. In a steeper gradient, the MLE method becomes more sensitive to uncertainties in cell positions. In this scenario, the tug-of-war model benefits from this limitation in the MLE method and senses the gradient more accurately, even in the presence of small positional errors. As a consequence, the trade-off values shift towards smaller positional uncertainties with the gradient steepness, see Fig. 4(d).

Our core results showing when clusters should prefer using tug-of-war or MLE are summarized in the phase diagram for the trade-off values shown in Fig. 4(e). In absence of positional error, the MLE method will always be preferred; the color map shows the amount of positional error required to make tug-of-war and MLE equivalent. We draw a dashed line at what we view as a typical positional uncertainty, $\Delta r_t = 2R_{\text{cell}}$. As we would expect from the scaling that $\sigma_g^2 \sim R_{\text{cluster}}^{-4}$ for MLE and $\sigma_g^2 \sim R_{\text{cluster}}^{-3}$ for tug-of-war, at a sufficiently large cluster size (large N), MLE will always be preferred – but the cluster size at which MLE is preferred depends strongly on the gradient strength.

Lastly, it is interesting to note the case of $N = 7$. This cluster size corresponds to the 1-layer hexagonal cluster, where almost all cells are edge cells, and there is only one interior cell. The interior cell is located at the cluster's center, so that $\delta\mathbf{r} = 0$, and therefore has no contribution to the gradient sensing in the MLE method (recall that $\chi = \sum_i^N |\delta\mathbf{r}_i|^2/2$ in the shallow gradient, see also for the general solution Eqs. (B4)–(B6)). Consequently, both models use the same cells to sense the gradient, except that the MLE approach is affected by positional uncertainty, whereas the tug-of-war is not. For this reason, even small uncertainties in cell positions are sufficient for the tug-of-war model to outperform the MLE and become the best gradient estimator. In the case of the shallow gradient approximation, we have that $\chi = \chi_{\text{tow}}$ and from Eq. (17) we obtain $\Delta r_t \approx 0$.

III. DISCUSSION

We have studied the role of limited positional information in collective gradient sensing. Within the context of gradient sensing, how cells get their positional information – or if they estimate the gradient without this information – is still an open question. Positional information within a group of cells has been studied in the context of developmental biology, where an external morphogen signal determine cell's fate [20, 22, 24, 41]; cells may also sense their position within a colony by sensing mechanical stresses [21]. Moving from a sensed chemical or mechanical signal to positional knowledge within a cluster is a complex multistage process [24]. Consequently, attaining precise positional information is a time costly process, and there are essential tradeoffs between the rate at which a pattern can be established and its precision [42]. Establishing a reproducible gradient is likely even more challenging for migratory cells in clusters [5, 13], as cell position and number of cells in the cluster all evolve over time. However, there is a potential candidate for a morphogen-like source of positional information within neural crest cell cluster chemotaxis: the complement fragment C3a, which acts as a “co-attractant” [43]. Earlier simulations show that the co-attractant C3a can be established in a graded fashion [44]. Here, we have suppressed the details of how positional information is established, neglecting potential issues with establishing a reliable gradient, and focused on the unavoidable impact of limited positional information in collective gradient sensing. Consequently, these results should in a sense be considered the best case for exploiting positional information: if the tradeoff argument of Fig. 4 predicts that maximum likelihood sensing using limited positional information is optimal, we should be skeptical about this. Contrarily, if the trade-off suggests MLE is inferior to the tug-of-war, we should be highly confident that this is true. For this reason, it seems unlikely that collective gradient sensing by small clusters, such as border cell migration [4, 45, 46], relies significantly on positional information.

Our results show that whether cells should use positional information in estimating a chemical gradient orientation depends not only on the amount of positional information but also on parameters like the cluster size and the gradient steepness. Large clusters overcome positional uncertainties by having numerous cells, *i.e.* more independent measurements, combining their estimates of their positions and the local concentration of chemoattractant. This large benefit from gaining information from all cells and weighting them according to position can outweigh the costs of having limited positional information. This is especially apparent in comparison with the tug-of-war model, which only uses information from the small fraction of the cells at the cluster's edge.

Gradient steepness is also a key parameter. Finite positional information limits the accuracy of collective gradient sensing by a group of cells – but this limitation

becomes more significant for steeper gradients, since difference in positions lead to larger changes in the chemoattractant concentration. Remarkably, for shallow gradients, ($g \rightarrow 0$), the limit where collective gradient sensing is most important, as single cells can become ineffective sensors, positional information turns out not to be relevant.

While our simplest analytical results are derived for a cluster with a constant positional information Δr , we can generalize our results beyond this point, finding that the effect of limited positional information differs depending on a cell's position within the cluster. Positional uncertainties in cells farther away from the cluster center of mass have a greater effect on the accuracy of gradient detection – again, suggesting that edge cells must play a large role in sensing. Our observations suggest that a group of cells may benefit from having specialized cells capable of sensing their position more accurately than the rest. The idea that cells may specialize depending on position in the cluster or chemoattractant concentration has also been suggested in other contexts [17, 34, 47, 48].

In addition to the maximum likelihood estimation of [19], which requires positional information, and the tug-of-war models [13, 16, 17], which do not, there are many other models of collective gradient sensing [15]. One broad category of a collective sensing mechanism that is not affected by positional information is collective local excitation–global inhibition (LEGI), where a local reporter reads out local chemoattractant concentration and a global inhibitor measures the global concentration [44, 49–52]. In LEGI, cell–cell communication imposes a maximum length scale at which gradient sensing information is reliably shared, leading to a saturation of gradient sensing accuracy with cell number [50], compatible with experiments on mammary organoids [14] but not lymphocyte clusters. While we have not treated the LEGI model in detail here, we argue that, similarly with the tug-of-war model, it does not require the cluster to know cell positions – but it does not gain the benefits of measuring signals over the entire cluster. We would similarly expect LEGI sensing to be suboptimal to the maximum likelihood estimator in the limit of increasingly large clusters, shallower gradients, or lowering positional uncertainty.

Can the consequences of limited positional information be characterized experimentally? In experiments tracking cluster migration in response to a chemoattractant gradient [5, 13], the distribution of the angle between the cluster's motion and the gradient is simple to measure, allowing σ_ϕ^2 to be measured. Earlier work on lymphocyte clusters studied cluster velocity and velocity variability as a function of cluster size to support a tug-of-war model in which chemotaxis was driven by forces along the edge, and noise in motility was driven by all cells, not just edge cells [13]; similar measurements have been made in the border cell cluster *in vivo* [53]. If a group of cells were reaching the optimal sensing limit given by Eq. (6), we would predict that accuracy σ_ϕ^2 would scale extremely

strongly with cluster size, $\sigma_\phi^2 \sim R_{\text{cluster}}^{-4}$ – a stronger dependence than that found by [13]. More characteristically, we would expect that as gradient strength g is increased, that angular accuracy σ_ϕ would saturate (Fig. 1d), in contrast with the predictions of previous models [19, 44], where increasing gradient strength essentially always increases cluster accuracy. However, this prediction should be interpreted cautiously: even in single cells, increasing gradient strength does not always continue to increase accuracy [54, 55]. This may reflect that at steep gradients, intracellular sources of noise in cell information processing and motility may be more important than at relatively shallow gradients, where the receptor-ligand noise is most important [55]. Within the shallow-gradient approximation, Eq. (6), σ_ϕ saturates to a value $\Delta r/\sqrt{\chi}$ at large gradient strength. The strongest evidence that positional information is playing a role as in Eq. (6) would be if, at large gradient strengths, this saturation value of σ_ϕ changed as sources of positional information were disrupted, altering Δr . This could arise from, e.g., interfering with the secretion or sensing of C3a in neural crest cells [43]. Ideal experiments would then extend earlier work [5] to a large range of cluster sizes with a more precise control over the chemical gradients clusters are exposed to. However, our earlier caveats on the model of Eq. (6) should be kept in mind: this result is a minimal model for positional error that is constant across the cluster, and more detailed comparisons would require more details in a putative source of positional information.

Our results show that a sensing mechanism that does not rely on positional information, such as tug-of-war, is optimal for sufficiently small clusters or sufficiently large gradients. In Figure 4, we have shown a phase diagram assuming that positional information is accurate to roughly a cell diameter, finding that tug-of-war is always optimal for clusters of seven cells, and will become optimal at large gradients for larger clusters. The experiments that originally suggested tug-of-war models where only edged cells participate in sensing and responding to the gradient [5, 13] involved tens to hundreds of cells. If the tug-of-war behavior represents an evolutionary optimum, it may reflect either a low ability to gain positional information or a typical need to follow relatively steep gradients.

ACKNOWLEDGEMENTS

BAC acknowledges support from NSF grant PHY 1915491. We thank Kurmanbek Kaiyrbekov for a close reading of the manuscript.

Appendix A: Computing the reading concentration error for a single cell

Cells estimate concentrations from the occupancy of receptors that binds the molecules of interest. Fluctua-

tions in the number of occupied receptors determine the accuracy of the cell to sense the concentration. To estimate the cell's sensing concentration error we first need to find how receptor occupancy fluctuates. We consider a cell with n_r independent receptors distributed along its surface and simple ligand-receptor kinetics, where molecules bind the receptors with a rate $k_{\text{on}}c(\mathbf{r}_i)$ and unbind with rate k_{off} . The probability that a receptor is bound to a ligand molecule is $p = \frac{k_{\text{on}}c(\mathbf{r}_i)}{k_{\text{on}}c(\mathbf{r}_i) + k_{\text{off}}} = \frac{c(\mathbf{r}_i)}{K_D + c(\mathbf{r}_i)}$, where $K_D = \frac{k_{\text{off}}}{k_{\text{on}}}$ is the reaction dissociation constant. Then, the mean number of occupied receptors is $\bar{n} = n_r \frac{c(\mathbf{r}_i)}{K_D + c(\mathbf{r}_i)}$. Since receptors can have only two states (bound or unbound), they follow a binomial process with variance $\delta n^2 = n_r p(1 - p) = \frac{n_r c(\mathbf{r}_i) K_D}{(K_D + c(\mathbf{r}_i))^2}$. Finally, the error in the concentration estimate is related to the fluctuations in the number of bound receptors through propagation-of-error, $\delta c^2 = \left(\frac{dc}{d\bar{n}}\right)^2 \delta n^2 = \frac{c(\mathbf{r}_i)(K_D + c(\mathbf{r}_i))^2}{n_r K_D}$.

Appendix B: Derivation of the Fisher Information matrix and computing of the lower bound for the gradient sensing error through the maximum likelihood estimation method

We apply MLE to obtain the gradient sensing error of a group of cells that sense their local chemoattractant concentration and have limited positional information. Each cell i performs two independent measurements, a measurement of the local concentration M_i and a measurement of its position \mathbf{r}_i^* given by Eqs. (1) and (2) respectively. From Section II A, we know we can write the likelihood function as, $\mathcal{L}(\mathbf{g}, \{\mathbf{r}_i\} | \{M_i, \mathbf{r}_i^*\}) = \prod_i^N p(M_i | \mathbf{g}, \mathbf{r}_i) p(\mathbf{r}_i^* | \mathbf{r}_i)$. In general, it is easier to work with the log of the likelihood instead, since the product between the probabilities becomes a sum. Then, considering a gradient $\mathbf{g} = g_x \hat{x} + g_y \hat{y}$ and a covariance matrix for the positional error,

$$\Sigma_i = \begin{bmatrix} \sigma_{x_i}^2 & \rho_i \sigma_{x_i} \sigma_{y_i} \\ \rho_i \sigma_{x_i} \sigma_{y_i} & \sigma_{y_i}^2 \end{bmatrix}, \quad (\text{B1})$$

we find the log of the likelihood to be

$$\begin{aligned} \log \mathcal{L}(\mathbf{g}, \{\mathbf{r}_i\} | \{M_i, \mathbf{r}_i^*\}) = & -\frac{3}{2} \sum_i^N \log(2\pi) - \frac{1}{2} \sum_i^N \log h_i - \sum_i^N \frac{(M_i - \mu_i)^2}{2h_i} - \frac{1}{2} \sum_i^N \log(\sigma_{x_i}^2 \sigma_{y_i}^2 (1 - \rho_i^2)) \\ & - \sum_i^N \frac{1}{2(1 - \rho_i^2)} \left[\frac{(r_{x_i}^* - r_{x_i})^2}{\sigma_{x_i}^2} - 2\rho_i \frac{(r_{x_i}^* - r_{x_i})(r_{y_i}^* - r_{y_i})}{\sigma_{x_i} \sigma_{y_i}} + \frac{(r_{y_i}^* - r_{y_i})^2}{\sigma_{y_i}^2} \right]. \end{aligned} \quad (\text{B2})$$

where $\mu_i = 1 + \mathbf{g} \cdot \delta \mathbf{r}_i$ and $h_i = (\delta c_i / c_0)^2 + \sigma_{\Delta}^2$ are the mean and variance of the local concentration measurement performed by cell i .

From Eq. (B2), we could find the global maximum to obtain the maximum likelihood estimator. Here, we are not interested in the estimator itself but in its fluctuations. We want to compute the estimator error, which we know must converge to the Cramér-Rao bound, given by the inverse of the Fisher information matrix, $(\mathcal{I})^{-1}$. This bound limits the accuracy of *any* unbiased measure-

ment of the gradient \mathbf{g} [30]. This bound puts a limit on minimal errors of any unbiased estimator of a parameter (e.g. an estimator $\hat{\alpha}$) away from the true value of that parameter (α). The Cramér-Rao bound for parameters α and β is,

$$\langle (\alpha - \hat{\alpha})(\beta - \hat{\beta}) \rangle = (\mathcal{I}^{-1})_{\alpha, \beta}, \quad (\text{B3})$$

Recalling the Fisher information definition, $\mathcal{I}_{\alpha, \beta} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial \alpha \partial \beta} \right\rangle$, we next take partial derivatives of Eq. (B2), and compute the expectation values, getting

$$\mathcal{I} = \begin{bmatrix} \sum_i^N f_i \delta r_{x_i}^2 & \sum_i^N f_i \delta r_{x_i} \delta r_{y_i} & g_x f_1 \delta r_{x_1} & g_y f_1 \delta r_{x_1} & \dots & g_x f_N \delta r_{x_N} & g_y f_N \delta r_{x_N} \\ \sum_i^N f_i \delta r_{x_i} \delta r_{y_i} & \sum_i^N f_i \delta r_{y_i}^2 & g_x f_1 \delta r_{y_1} & g_y f_1 \delta r_{y_1} & \dots & g_x f_N \delta r_{y_N} & g_y f_N \delta r_{y_N} \\ g_x f_1 \delta r_{x_1} & g_x f_1 \delta r_{y_1} & g_x^2 f_1 + \frac{m_1}{\sigma_{x_1}^2} & g_x g_y f_1 - \frac{m_1 \rho_1}{\sigma_{x_1} \sigma_{y_1}} & \dots & 0 & 0 \\ g_y f_1 \delta r_{x_1} & g_y f_1 \delta r_{y_1} & g_x g_y f_1 - \frac{m_1 \rho_1}{\sigma_{x_1} \sigma_{y_1}} & g_y^2 f_1 + \frac{m_1}{\sigma_{y_1}^2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ g_x f_N \delta r_{x_N} & g_x f_N \delta r_{y_N} & 0 & 0 & \dots & g_x^2 f_N + \frac{m_N}{\sigma_{x_N}^2} & g_x g_y f_N - \frac{m_N \rho_N}{\sigma_{x_N} \sigma_{y_N}} \\ g_y f_N \delta r_{x_N} & g_y f_N \delta r_{y_N} & 0 & 0 & \dots & g_x g_y f_N - \frac{m_N \rho_N}{\sigma_{x_N} \sigma_{y_N}} & g_y^2 f_N + \frac{m_N}{\sigma_{y_N}^2} \end{bmatrix}$$

where $f_i = \frac{(a+\mu_i)^2((a+3\mu_i)^2+2an_r\mu_i)+2a^2n_r^2\sigma_\Delta^2}{2(\mu_i(\mu_i+a)^2+an_r\sigma_\Delta^2)^2}$, $m_i = \frac{1}{(1-\rho_i^2)}$, $a = K_D/c_0$, and $\delta\mathbf{r}_i = \mathbf{r}_i - \mathbf{r}_{\text{cm}}$. The order of the Fisher information matrix elements correspond to: $\{g_x, g_y, r_{x_1}, r_{y_1}, \dots, r_{x_N}, r_{y_N}\}$, such that, as for example, $\mathcal{I}_{1,1} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial g_x^2} \right\rangle$, $\mathcal{I}_{1,2} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial g_x \partial g_y} \right\rangle$, $\mathcal{I}_{1,2(N+1)} = -\left\langle \frac{\partial^2 \ln \mathcal{L}}{\partial g_x \partial r_{y_N}} \right\rangle$. The derivation of the Fisher information matrix from the likelihood function is not hard, but takes some calculation effort. To gain some intuition about the shape it takes, we point out some aspects. Note that \mathbf{g} and \mathbf{r}_i are only related by the likelihood function through the term $\mathbf{g} \cdot \delta\mathbf{r}_i$, which appears inside the expressions of h_i and μ_i . From there we can see why the f_i parameters are in every element of the Fisher information matrix. Taking second derivatives with respect to \mathbf{g} and \mathbf{r}_i over the different terms in Eq. (B2), by the derivative rule of chain, results in ending up taking derivatives over $\mathbf{g} \cdot \delta\mathbf{r}_i$. Then, all these terms factor together yielding to f_i . For the same reason we can understand \mathcal{I} having the first two rows/columns in full. Lastly, note that the elements related to the double derivatives with respect to positions have a contribution coming from the gradient terms, plus the contribution from the positional uncertainty terms of the likelihood function.

We are interested in computing the gradient sensing errors, $\sigma_{g_x}^2 \equiv \langle (g_x - \hat{g}_x)^2 \rangle = (\mathcal{I}^{-1})_{g_x, g_x}$ and $\sigma_{g_y}^2 \equiv \langle (g_y - \hat{g}_y)^2 \rangle = (\mathcal{I}^{-1})_{g_y, g_y}$. Following a guided Gaussian elimination procedure we can obtain the inverse of the Fisher information matrix for the elements associated with the gradient (see Appendix E). The maximum likelihood estimator errors for the gradient are,

$$\sigma_{g_x}^2 = \frac{1}{\mathcal{S}} \sum_i^N \gamma_i \delta r_{y_i}^2, \quad (\text{B4})$$

$$\sigma_{g_y}^2 = \frac{1}{\mathcal{S}} \sum_i^N \gamma_i \delta r_{x_i}^2, \quad (\text{B5})$$

$$\sigma_{g_x, g_y} = -\frac{1}{\mathcal{S}} \sum_i^N \gamma_i \delta r_{x_i} \delta r_{y_i}, \quad (\text{B6})$$

where $\mathcal{S} = \sum_i^N \gamma_i \delta r_{x_i}^2 \sum_i^N \gamma_i \delta r_{y_i}^2 - \left(\sum_i^N \gamma_i \delta r_{x_i} \delta r_{y_i} \right)^2$, and $\gamma_i = \frac{f_i}{1+f_i \mathbf{g}^T \boldsymbol{\Sigma}_i \mathbf{g}}$.

Finally, it is easier to interpret the gradient sensing errors in terms of the steepness, g , and direction ϕ , of the gradient. We can obtain them by re-parametrizing the Fisher information matrix as,

$$I_{\theta'} = \mathbf{J}^T I_\theta(\theta(\theta')) \mathbf{J},$$

where \mathbf{J} is the Jacobian matrix, $J_{ij} = \frac{\partial \theta_i}{\partial \theta_j}$, and $\theta = \{g_x, g_y\}$ and $\theta' = \{g, \phi\}$ are the variables we want to transform. The errors for the gradient steepness and di-

rection then are,

$$\sigma_g^2 = \cos^2(\phi) \sigma_{g_x}^2 + \sin^2(\phi) \sigma_{g_y}^2 + 2 \cos(\phi) \sin(\phi) \sigma_{g_x, g_y} \quad (\text{B7})$$

$$\sigma_\phi^2 = \frac{\sin^2(\phi) \sigma_{g_x}^2 + \cos^2(\phi) \sigma_{g_y}^2 - 2 \cos(\phi) \sin(\phi) \sigma_{g_x, g_y}}{g^2}. \quad (\text{B8})$$

We presented these equations for an isotropic cluster in the main text, where $\sigma_{g_x, g_y} \approx 0$ and $\sigma_{g_x} \approx \sigma_{g_y}$. In this case, we find that $\sigma_\phi^2 = \sigma_g^2/g^2$ – but this is only true for isotropic clusters. Note that Eq. (B8) fails for large enough σ_g , since ϕ is constrained between $0-2\pi$, and consequently, σ_ϕ is bounded.

Shallow gradient approximation

In the limit of shallow gradient approximation it is possible to obtain simpler expression for the gradient sensing errors. In this limit, $gR_{\text{cluster}} \ll 1$, meaning that the cell's ligand-receptor fluctuations can be approximated as $\delta c_i \approx \bar{c} = \sqrt{\frac{1}{n_r} \frac{c_0(c_0+K_D)^2}{K_D}}$. Variations of the chemoattractant concentration along the cell's cluster are small, thus we can assume that all cells have the same reading concentration errors. Then, $h_i \approx \bar{h} = \bar{c}^2 + \sigma_\Delta^2$. Moreover, since $n_r \gg K_D/c_0$, f_i can be approximated as the inverse of the reading concentration error, $f_i \approx \bar{f} = 1/\bar{h}$. Note that the limit of large cell-cell variability, $\sigma_\Delta \gg \delta c_i/c_0$, would lead to the same approximation. Assuming constant isotropic positional errors, such that $\rho_i = 0$ and $\sigma_{x_i} = \sigma_{y_i} = \Delta r$, we find that $\gamma_i = \frac{1}{1/f_i + \mathbf{g}^T \boldsymbol{\Sigma}_i \mathbf{g}} \approx \frac{1}{\bar{h} + g^2 \Delta r^2}$ is independent of i and thus Eqs. (B4)–(B6) take a simplified form,

$$\sigma_{g_x}^2 = \frac{\sum_i^N \delta r_{y_i}^2}{\bar{\mathcal{S}}} (\bar{h} + g^2 \Delta r^2),$$

$$\sigma_{g_y}^2 = \frac{\sum_i^N \delta r_{x_i}^2}{\bar{\mathcal{S}}} (\bar{h} + g^2 \Delta r^2),$$

and in polar coordinates,

$$\sigma_g^2 = \Gamma_g (\bar{h} + g^2 \Delta r^2),$$

$$\sigma_\phi^2 = \frac{\Gamma_\phi}{g^2} (\bar{h} + g^2 \Delta r^2),$$

where $\Gamma_g = \frac{\sum_i^N (\sin(\phi) \delta r_{x_i} - \cos(\phi) \delta r_{y_i})^2}{\bar{\mathcal{S}}}$, $\Gamma_\phi = \frac{\sum_i^N (\cos(\phi) \delta r_{x_i} + \sin(\phi) \delta r_{y_i})^2}{\bar{\mathcal{S}}}$, and $\bar{\mathcal{S}} = \sum_i^N \delta r_{x_i}^2 \sum_i^N \delta r_{y_i}^2 - (\sum_i^N \delta r_{x_i} \delta r_{y_i})^2$ are geometrical factors that depend on the cluster's shape and the gradient's orientation. Note that in the case the cluster have circular symmetry, $(\sum_i^N \delta r_{x_i}^2 \approx \sum_i^N \delta r_{y_i}^2 \approx \frac{1}{2} \sum_i^N |\delta \mathbf{r}_i|^2, \sum_i^N \delta r_{x_i} \delta r_{y_i} \approx 0)$, then $\Gamma_g \approx \frac{1}{\chi}$, and Eq. (6) is recovered.

Appendix C: Auxiliary computations for the tug-of-war model

Details on computation of the mean and covariance of $\mathbf{p}_i(t)$

We want to compute the mean value and variance of \mathbf{p}_i . We start from Eq. (12). The mean value is straight forward,

$$\begin{aligned}\langle \mathbf{p}_i \rangle &= \left\langle \boldsymbol{\mu}_i + \sigma_p \int_0^t dt' e^{-(t-t')/\tau_p} \boldsymbol{\epsilon}_i(t') + \bar{\beta} c_0 \mathbf{q}_i \int_0^t dt' e^{-(t-t')/\tau_p} \Xi_i(t') \right\rangle, \\ &= \langle \boldsymbol{\mu}_i \rangle + \sigma_p \int_0^t dt' e^{-(t-t')/\tau_p} \langle \boldsymbol{\epsilon}_i(t') \rangle + \bar{\beta} c_0 \mathbf{q}_i \int_0^t dt' e^{-(t-t')/\tau_p} \langle \Xi_i(t') \rangle, \\ &= \boldsymbol{\mu}_i.\end{aligned}$$

Recalling that $\langle \Xi_i(t) \Xi_j(0) \rangle = (\delta c_i^2/c_0^2 C_c(t) + \sigma_\Delta^2 C_\Delta(t)) \delta_{ij}$, and that we have assumed that correlations decay exponentially, so $C_s(t) = e^{-t/\tau_s}$,

$$\begin{aligned}\langle (p_{\mu_i}(t) - \langle p_{\mu_i} \rangle)(p_{\nu_i}(t) - \langle p_{\nu_i} \rangle) \rangle &= \sigma^2 \int_0^t dt' \int_0^t dt'' \langle \epsilon_{\mu_i}(t') \epsilon_{\nu_i}(t'') \rangle e^{-(t-t')/\tau_p} e^{-(t-t'')/\tau_p} \\ &\quad + \bar{\beta}^2 c_0^2 q_{\mu_i} q_{\nu_i} \int_0^t dt' \int_0^t dt'' \langle \Xi_i(t') \Xi_i(t'') \rangle e^{-(t-t')/\tau_p} e^{-(t-t'')/\tau_p} \\ &= \frac{\sigma^2 \tau_p}{2} + \bar{\beta}^2 q_{\mu_i} q_{\nu_i} \delta c_i^2 T(t, \tau_c) + \bar{\beta}^2 c_0^2 q_{\mu_i} q_{\nu_i} \sigma_\Delta^2 T(t, \tau_\Delta)\end{aligned}$$

where,

$$\begin{aligned}T(t, \tau_s) &= \frac{\tau_p^2}{1 + v_s} \left(1 + \frac{1}{1 - v_s} \right. \\ &\quad \times \left. ((1 + v_s) e^{-2t/\tau_p} - 2e^{-(1+v_s)t/\tau_p}) \right),\end{aligned}$$

and $v_s = \frac{\tau_p}{\tau_s}$. $T(t, \tau_s)/\tau_p^2$ represent the time averaging function for a fluctuating process with correlation time τ_s up to a time t . Given that $\tau_\Delta \gg t \gg \tau_p \gg \tau_c$, then $T(t, \tau_c) \approx \frac{\tau_p^2}{1 + \tau_p/\tau_c} \approx \tau_p \tau_c$ and $T(t, \tau_\Delta) \approx \tau_p^2$, so then $T(t, \tau_\Delta) \gg T(t, \tau_c)$.

Error propagation for the tug-of-war model

To estimate the gradient sensing error for the tug-of-war model we use a simple error propagation method.

with τ_s the correlation time and the index $s = \{c, \Delta\}$, we now compute the covariance of $\mathbf{p}_i(t)$,

From Eq. (13), we can write an expression for the gradient related to the cluster velocity, $\mathbf{g} = \frac{1}{\gamma c_0} \mathcal{M}^{-1} \langle \mathbf{v}_c \rangle$, where,

$$\mathcal{M}^{-1} = \frac{N}{\mathcal{S}_\mathbf{v}} \begin{bmatrix} \sum_i^N \delta r_{y_i} q_{y_i} & -\sum_i^N \delta r_{y_i} q_{x_i} \\ -\sum_i^N \delta r_{x_i} q_{y_i} & \sum_i^N \delta r_{x_i} q_{x_i} \end{bmatrix}, \quad (C1)$$

$$\text{and } \mathcal{S}_\mathbf{v} = \sum_i^N \delta r_{x_i} q_{x_i} \sum_i^N \delta r_{y_i} q_{y_i} - \sum_i^N \delta r_{x_i} q_{y_i} \sum_i^N \delta r_{y_i} q_{x_i}.$$

Recalling the symmetry of the roughly circular hexagonal clusters, the following relations are fulfilled, $\sum_i^N \delta r_{x_i}^2 \approx \sum_i^N \delta r_{y_i}^2 \approx \frac{1}{2} \sum_i^N |\mathbf{r}_i|^2$, $\sum_i^N \delta r_{x_i} \delta r_{y_i} \approx 0$, $\sum_i^N q_{x_i}^2 \approx \sum_i^N q_{y_i}^2 \approx \frac{1}{2} \sum_i^N |\mathbf{q}_i|^2$, $\sum_i^N q_{x_i} q_{y_i} \approx 0$, $\sum_i^N q_{x_i} \delta r_{x_i} \approx \sum_i^N q_{y_i} \delta r_{y_i} \approx \frac{1}{2} \sum_i^N \mathbf{q}_i \cdot \delta \mathbf{r}_i$, and $\sum_i^N q_{x_i} \delta r_{y_i} \approx \sum_i^N q_{y_i} \delta r_{x_i} \approx 0$. In addition, note that in such a case, we can write $\mathcal{S}_\mathbf{v} \approx \sum_i^N \delta r_{x_i} q_{x_i} \sum_i^N \delta r_{y_i} q_{y_i} \approx$

$\frac{1}{4} \left(\sum_i^N \mathbf{q}_i \cdot \delta \mathbf{r}_i \right)^2$ and,

$$\mathcal{M}^{-1} \approx \frac{2N}{\sum_i^N \mathbf{q}_i \cdot \delta \mathbf{r}_i} \mathbb{I}, \quad (\text{C2})$$

where \mathbb{I} is the 2×2 identity matrix. Then, the expression for the gradient takes the following simple form,

$$\mathbf{g} = \frac{2N}{\gamma c_0 \sum_i^N \mathbf{q}_i \cdot \delta \mathbf{r}_i} \langle \mathbf{v}_c \rangle. \quad (\text{C3})$$

Eq. (C3) shows that \mathbf{g} and $\langle \mathbf{v}_c \rangle$ are proportional to each other. Then, we can write the variance of \mathbf{g} as,

$$\text{var}(\mathbf{g}) = \left(\frac{2N}{\gamma c_0 \sum_i^N \mathbf{q}_i \cdot \delta \mathbf{r}_i} \right)^2 \text{var}(\mathbf{v}_c). \quad (\text{C4})$$

Introducing the velocity variance expressions given by Eq. (14) in the main text into Eq. (C4), we finally obtain the gradient sensing variances for the tug-of-war model,

$$\sigma_{g_x}^2 \approx \sigma_{g_y}^2 \approx \frac{2}{\left(\sum_i^N \mathbf{q}_i \cdot \delta \mathbf{r}_i \right)^2} \left(\sum_i^N |\mathbf{q}_i|^2 h_{T_i} + \frac{\sigma_p^2 \tau_p}{\gamma^2 c_0^2} N \right), \quad (\text{C5})$$

$$\text{cov}(g_x, g_y) \approx 0. \quad (\text{C6})$$

We note that, though our derivation of this equation does involve a sum over cell positions, the cluster can compute the direction of the gradient through the tug-of-war *without* explicitly knowing any cell locations.

Appendix D: Maximum likelihood estimator with time averaging

In the main text we apply the MLE method to compute the gradient sensing error from the individual cell

measurements of the chemoattractant concentration, M_i , and cell positions, \mathbf{r}_i^* . These measurements capture a “snapshot” or instant of the cluster state. Later in the text, we compare the MLE outcomes with the tug-of-war model, which relies on the statistics of the cell polarities \mathbf{p}_i obtained for asymptotic times larger than τ_p . Therefore, we are comparing instantaneous with time averaged measurements. We could do this since the sensing concentration measurements are dominated by the CCV which has a correlation times larger than the polarity relaxation time. Here we extend the MLE method by using time average measurements in the same way as the tug-of-war model.

We compute the time average for the individual cell chemoattractant concentration measurements,

$$M_i^T(t) = \int M_i(t') K_T(t - t') dt',$$

$$= c(\mathbf{r}_i)/c_0 + \int \Xi_i(t') K_T(t - t') dt',$$

where $K_T(t) = \mathcal{H}(t) \frac{1}{T} e^{-t/T}$ is a time averaging kernel and $\mathcal{H}(t)$ the Heaviside function. We assume that cell positions remain fixed during the averaging time T , so that $c(\mathbf{r}_i, t) = c(\mathbf{r}_i)$. Now we compute the mean and variance for M_i^T .

$$\langle M_i^T \rangle = c(\mathbf{r}_i)/c_0,$$

$$\begin{aligned} \langle (M_i^T - \langle M_i^T \rangle)^2 \rangle &= \frac{1}{T^2} \int_0^t dt' \int_0^t dt'' \langle \Xi_i(t') \Xi_i(t'') \rangle e^{-(t-t')/T} e^{-(t-t'')/T}, \\ &= \frac{1}{T^2} \int_0^t dt' \int_0^t dt'' \left(\frac{\delta c_i^2}{c_0^2} C_c(|t' - t''|) + \sigma_\Delta^2 C_\Delta(|t' - t''|) \right) e^{-(t-t')/T} e^{-(t-t'')/T}, \\ &= \frac{1}{1 + T/\tau_c} \frac{\delta c_i^2}{c_0^2} + \frac{1}{1 + T/\tau_\Delta} \sigma_\Delta^2. \end{aligned}$$

If we substitute the averaging time $T = \tau_p$, these are precisely the values h_{T_i} which we derived in our tug-of-war model. This shows we can carry over all of our results from the maximum likelihood estimation to match the results of the tug-of-war model. However, this is not

particularly important, because $T \gg \tau_c$ and $T \ll \tau_\Delta$, the sensing concentration fluctuations can be approximated by $\langle (M_i^T - \langle M_i^T \rangle)^2 \rangle \approx \sigma_\Delta^2$. When we take into account time averaging, the contributions to the sensing concentration fluctuations are dominated by CCV while

ligand-receptor fluctuations are averaged out.

Appendix E: Proof for the analytical solution of the MLE errors

In this appendix we show the derivation of the expressions for the MLE errors in the shallow gradient limit. The problem we are solving is very similar to an error-in-variable models of regression, in which the regressor variable is subject to errors [56]. However, we were not able to find an explicit mapping between these results and our question of interest – computing the Fisher information matrix and its inverse specifically for the g_x and g_y variables. Finding the Fisher information matrix \mathcal{I} is relatively straightforward from evaluating derivatives of the log-likelihood, and it is presented in Appendix B. However, to determine the Cramer-Rao bound, we need to compute the *inverse* of the Fisher information matrix – in principle needing to analytically invert a matrix

whose size scales with the number of cells. This is non-trivial in general. However, for this problem we were able to use Gaussian elimination to find the elements of the inverse of the Fisher information matrix necessary for us, *i.e.* $(\mathcal{I})_{g_x g_x}^{-1}$, $(\mathcal{I})_{g_y g_y}^{-1}$, and $(\mathcal{I})_{g_x g_y}^{-1}$. Given the almost tridiagonal shape of the Fisher matrix, we first show the procedure for how to eliminate the elements of the rows associated with a individual cell s and then repeat the same steps for the rest of the cells.

Recall that the Gaussian elimination method is a linear algebra method to find the solution of a problem $\mathcal{A} \times X = \mathcal{B}$. This problem can be represented in terms of an augmented matrix $[\mathcal{A}|\mathcal{B}]$. Row operations (*i.e.* linear combinations of the rows) are performed, leading eventually to finding the solution for X when the augmented matrix takes the form $[\mathbb{I}|\mathcal{C}]$, where \mathbb{I} is the identity matrix, and $\mathcal{C} = (\mathcal{A}^{-1}) \mathcal{B} = X$. In our case, the problem we want to solve is to find the inverse of the Fisher information matrix, $\mathcal{I} \times (\mathcal{I}^{-1}) = \mathbb{I}$. To start with the Gaussian elimination process, we first write the augmented matrix $[\mathcal{I}|\mathbb{I}]$,

$$\left[\begin{array}{cccccc|ccc} S_{xx} & S_{xy} & \dots & g_x f_s \delta r_{x_s} & g_y f_s \delta r_{x_s} & \dots & 1 & 0 & \dots \\ S_{xy} & S_{yy} & \dots & g_x f_s \delta r_{y_s} & g_y f_s \delta r_{y_s} & \dots & 0 & 1 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \\ g_x f_s \delta r_{x_s} & g_x f_s \delta r_{y_s} & \dots & g_x^2 f_s + \frac{m_s}{\sigma_{x_s}^2} & g_x g_y f_s + \frac{m_s \rho_s}{\sigma_{x_s} \sigma_{y_s}} & \dots & 0 & 0 & \dots \\ g_y f_s \delta r_{x_s} & g_y f_s \delta r_{y_s} & \dots & g_x g_y f_s + \frac{m_s \rho_s}{\sigma_{x_s} \sigma_{y_s}} & g_y^2 f_s + \frac{m_s}{\sigma_{y_s}^2} & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \end{array} \right]$$

where $S_{xx} = \sum_i^N f_i \delta r_{x_i}^2$, $S_{yy} = \sum_i^N f_i \delta r_{y_i}^2$, and $S_{xy} = \sum_i^N f_i \delta r_{x_i} \delta r_{y_i}$. Since we are only looking for the gradient estimation errors, we solely consider the two first columns of the right matrix.

Following, we apply the operations to perform an upper diagonalization of columns associated to the cell s . From now on we use the notation R^i and $\mathcal{I}_{m,n}$ to identify the i -row and the (m,n) -element of the matrix respectively. Note that the rows associated with the cell s are R^{2s+1} and R^{2s+2} .

1. Set element $\mathcal{I}_{2s+2,2s+2} = 1$: $R^{2s+2} \leftarrow \frac{R^{2s+2}}{\mathcal{I}_{2s+2,2s+2}}$.
2. Set element $\mathcal{I}_{2s+1,2s+2} = 0$: $R^{2s+1} \leftarrow R^{2s+1} - \mathcal{I}_{2s+1,2s+2} R^{2s+2}$.
3. Set element $\mathcal{I}_{2,2s+2} = 0$: $R^2 \leftarrow R^2 - \mathcal{I}_{2,2s+2} R^{2s+2}$.
4. Set element $\mathcal{I}_{1,2s+2} = 0$: $R^1 \leftarrow R^1 - \mathcal{I}_{1,2s+2} R^{2s+2}$.
5. Set element $\mathcal{I}_{2s+1,2s+1} = 1$: $R^{2s+1} \leftarrow \frac{R^{2s+1}}{\mathcal{I}_{2s+1,2s+1}}$.
6. Set element $\mathcal{I}_{2,2s+1} = 0$: $R^2 \leftarrow R^2 - \mathcal{I}_{2,2s+1} R^{2s+1}$.
7. Set element $\mathcal{I}_{1,2s+1} = 0$: $R^1 \leftarrow R^1 - \mathcal{I}_{1,2s+1} R^{2s+1}$.

After these steps we arrive to the following matrix,

$$\left[\begin{array}{cccccc|ccc} S_{xx} - \alpha_s \delta r_{x_s}^2 & S_{xy} - \alpha_s \delta r_{x_s} \delta r_{y_s} & \dots & 0 & 0 & \dots & 1 & 0 & \dots \\ S_{xy} - \alpha_s \delta r_{x_s} \delta r_{y_s} & S_{yy} - \alpha_s \delta r_{y_s}^2 & \dots & 0 & 0 & \dots & 0 & 1 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \\ \square & \square & \dots & 1 & 0 & \dots & 0 & 0 & \dots \\ \square & \square & \dots & \square & 1 & \dots & 0 & 0 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots & \vdots & \vdots & \dots \end{array} \right]$$

$\alpha_s = f_s^2 \frac{\mathbf{g} \mathbf{\Sigma}_s \mathbf{g}^T}{1 + f_s \mathbf{g} \mathbf{\Sigma}_s \mathbf{g}^T}$ and the squares (\square) represent elements with large expressions that are not relevant for the computation.

Then, we need to repeat steps 1–7 for all s . For each iteration we add a term $-\alpha_s \delta r_{z_k s} \delta r_{z_l s}$ to the matrix elements $\{k,l\}$ with $k,l = \{1,2\}$ and $z_{k,l} = \{x,y\}$. Finally, after repeating the procedure for all s , we get that the elements $\{k,l\}$ results in $\sum_i^N f_i \delta r_{z_x i} \delta r_{z_l i} - \sum_s \alpha_s \delta r_{z_x s} \delta r_{z_l s} = \sum_i^N \gamma_i \delta r_{z_x i} \delta r_{z_l i}$,

where $\gamma_i = \frac{f_i}{1+f_i\mathbf{g}^T\boldsymbol{\Sigma}_i\mathbf{g}^T}$, and the matrix reads,

$$\left[\begin{array}{ccc|ccc} \sum_i^N \gamma_i \delta r_{x_i}^2 & \sum_i^N \gamma_i \delta r_{x_i} \delta r_{y_i} & \dots & 1 & 0 & \dots \\ \sum_i^N \gamma_i \delta r_{x_i} \delta r_{y_i} & \sum_i^N \gamma_i \delta r_{y_i}^2 & \dots & 0 & 1 & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots \end{array} \right]$$

Finally, since the first two rows are composed by a 2×2 block followed by all zeros to the right, instead of performing row operations, we can just invert this 2×2 matrix block to arrive to the final solution for the inverse of the Fisher matrix and the MLE errors.

$$\left[\begin{array}{c|ccc} 1 & 0 & \dots & \frac{\sum_i^N \gamma_i \delta r_{y_i}^2}{\mathcal{S}} & -\frac{\sum_i^N \gamma_i \delta r_{x_i} \delta r_{y_i}}{\mathcal{S}} & \dots \\ 0 & 1 & \dots & -\frac{\sum_i^N \gamma_i \delta r_{x_i} \delta r_{y_i}}{\mathcal{S}} & \frac{\sum_i^N \gamma_i \delta r_{x_i}^2}{\mathcal{S}} & \dots \\ \vdots & \vdots & \dots & \vdots & \vdots & \dots \end{array} \right]$$

We can be positive that this is the final solution, since following with the Gaussian elimination process, no further operations over the first two rows would be necessary, and therefore, they will not be modified.

We have ensured that this result is correct both by checking with computer algebra packages for small numbers of cells, as well as explicitly numerically inverting the Fisher information matrix.

Appendix F: Default parameters

In Table I we show the default values for parameters in this study with a brief justification of our choices. Values differing from these are indicated in the text and the figure captions.

parameters	value	justification
Q	3	$N = 8 - 100$. From [13] area of clusters is $\sim 2500\mu\text{m}^2$ and considering that $R_{\text{cell}} \sim 3 - 10\mu\text{m}$.
R_{cell}	$10\mu\text{m}$	Lymphocytes $\sim 3 - 5\mu\text{m}$, epithelial cells $\sim 4 - 10\mu\text{m}$. From [57, 58].
n_r	10^5	
$a = \frac{K_d}{c_0}$	1	Optimal setting for reducing ligand-receptor noise [19].
Δr	$2R_{\text{cell}}$ (1 cell)	Error estimation in <i>Drosophila m.</i> embryos [20]
g_0	$0.005\mu\text{m}^{-1}$	According to [19] and [13].
σ_Δ	0.1	According to [59]
σ_p	0	No polarization error. Best estimation the tug-of-war model can make.

TABLE I. Parameter selection. Unless otherwise indicated in the text, these are the parameters used throughout the results.

REFERENCES

- [1] Shuvasree SenGupta, Carole A Parent, and James E Bear. The principles of directed cell migration. *Nature Reviews Molecular Cell Biology*, pages 1–19, 2021.
- [2] Herbert Levine and Wouter-Jan Rappel. The physics of eukaryotic chemotaxis. *Physics Today*, 66(2), 2013.
- [3] Philippe V Afonso, Mirkka Janka-Juntila, Young Jong Lee, Colin P McCann, Charlotte M Oliver, Khaled A Aamer, Wolfgang Losert, Marcus T Cicerone, and Carole A Parent. LTB4 is a signal-relay molecule during neutrophil chemotaxis. *Developmental cell*, 22(5):1079–1091, 2012.
- [4] Denise J Montell, Wan Hee Yoon, and Michelle Starz-Gaiano. Group choreography: mechanisms orchestrating the collective movement of border cells. *Nature Reviews Molecular Cell Biology*, 13(10):631–645, 2012.
- [5] Eric Thveneau, Lorena Marchant, Sei Kuriyama, Mazhar Gull, Barbara Moepps, Maddy Parsons, and Roberto Mayor. Collective chemotaxis requires contact-dependent cell polarity. *Developmental cell*, 19(1):39–53, 2010.
- [6] Christina H Stuelten, Carole A Parent, and Denise J Montell. Cell motility in cancer invasion and metastasis: insights from simple model organisms. *Nature Reviews Cancer*, 18(5):296–312, 2018.
- [7] Evangelia T Roussos, John S Condeelis, and Antonia Patsialou. Chemotaxis in cancer. *Nature Reviews Cancer*, 11(8):573–587, 2011.
- [8] Abhinava K Mishra, Joseph P Campanale, James A Mondo, and Denise J Montell. Cell interactions in collective cell migration. *Development*, 146(23), 2019.
- [9] Peter Friedl and Darren Gilmour. Collective cell migration in morphogenesis, regeneration and cancer. *Nature Reviews Molecular Cell Biology*, 10(7):445–457, 2009.
- [10] Anna Haeger, Katarina Wolf, Mirjam M Zegers, and Peter Friedl. Collective cell migration: guidance principles and hierarchies. *Trends in Cell Biology*, 25(9):556–566, 2015.
- [11] Kevin J Cheung and Andrew J Ewald. A collective route to metastasis: Seeding by tumor cell clusters. *Science*, 352(6282):167–169, 2016.
- [12] Roberto Mayor and Sandrine Etienne-Manneville. The front and rear of collective cell migration. *Nature reviews Molecular cell biology*, 17(2):97, 2016.
- [13] Gema Malet-Engra, Weimiao Yu, Amanda Oldani, Javier Rey-Barroso, Nir S Gov, Giorgio Scita, and Loïc Dupré. Collective cell motility promotes chemotactic prowess and resistance to chemorepulsion. *Current Biology*, 25(2):242–250, 2015.

[14] David Ellison, Andrew Mugler, Matthew D Brennan, Sung Hoon Lee, Robert J Huebner, Eliah R Shamir, Laura A Woo, Joseph Kim, Patrick Amar, Ilya Nemenman, et al. Cell-cell communication enhances the capacity of cell ensembles to sense shallow gradients during morphogenesis. *Proceedings of the National Academy of Sciences*, 113(6):E679–E688, 2016.

[15] Brian A Camley. Collective gradient sensing and chemotaxis: modeling and recent developments. *Journal of Physics: Condensed Matter*, 30(22):223001, 2018.

[16] Brian A Camley, Juliane Zimmermann, Herbert Levine, and Wouter-Jan Rappel. Emergent collective chemotaxis without single-cell gradient sensing. *Physical Review Letters*, 116(9):098101, 2016.

[17] Katherine Copenhagen, Gema Malet-Engra, Weimiao Yu, Giorgio Scita, Nir Gov, and Ajay Gopinathan. Frustration-induced phases in migrating cell clusters. *Science Advances*, 4(9):eaar8483, 2018.

[18] This assumes that interior cells can sense the concentration. In three-dimensional organoids, interior cells are isolated and may not have an equal ability to measure concentration.

[19] Brian A Camley and Wouter-Jan Rappel. Cell-to-cell variation sets a tissue-rheology-dependent bound on collective gradient sensing. *Proceedings of the National Academy of Sciences*, 114(47):E10074–E10082, 2017.

[20] Thomas Gregor, David W Tank, Eric F Wieschaus, and William Bialek. Probing the limits to positional information. *Cell*, 130(1):153–164, 2007.

[21] Hayden Nunley, Xufeng Xue, Jianping Fu, and David K Lubensky. Generation of fate patterns via intercellular forces. *bioRxiv*, 2021.

[22] Lewis Wolpert. Positional information revisited. *Development*, 107(Supplement):3–12, 1989.

[23] Lewis Wolpert. Positional information and patterning revisited. *Journal of Theoretical Biology*, 269(1):359–365, 2011.

[24] Gašper Tkačík and Thomas Gregor. The many bits of positional information. *Development*, 148(2):dev176065, 2021.

[25] Julien O Dubuis, Gašper Tkačík, Eric F Wieschaus, Thomas Gregor, and William Bialek. Positional information, in bits. *Proceedings of the National Academy of Sciences*, 110(41):16301–16308, 2013.

[26] Emiliano Pérez Ipiña and Silvina Ponce Dawson. Fluctuations, correlations and the estimation of concentrations inside cells. *PloS ONE*, 11(3), 2016.

[27] Howard C Berg and Edward M Purcell. Physics of chemoreception. *Biophysical Journal*, 20(2):193–219, 1977.

[28] William Bialek and Sima Setayeshgar. Physical limits to biochemical signaling. *Proceedings of the National Academy of Sciences*, 102(29):10040–10045, 2005.

[29] Kazunari Kaizu, Wiet De Ronde, Joris Pajjmans, Koichi Takahashi, Filipe Tostevin, and Pieter Rein Ten Wolde. The berg-purcell limit revisited. *Biophysical Journal*, 106(4):976–985, 2014.

[30] Steven M Kay. *Fundamentals of statistical signal processing, Volume I: Estimation Theory*. Pearson, 1993.

[31] Bo Hu, Wen Chen, Wouter-Jan Rappel, and Herbert Levine. Physical limits on cellular sensing of spatial gradients. *Physical Review Letters*, 105(4):048104, 2010.

[32] Bo Hu, Wen Chen, Wouter-Jan Rappel, and Herbert Levine. How geometry and internal bias affect the accuracy of eukaryotic gradient sensing. *Physical Review E*, 83(2):021917, 2011.

[33] Robert G Endres and Ned S Wingreen. Maximum likelihood and the single receptor. *Physical Review Letters*, 103(15):158101, 2009.

[34] Austin Hopkins and Brian A Camley. Chemotaxis in uncertain environments: Hedging bets with multiple receptor types. *Physical Review Research*, 2(4):043146, 2020.

[35] Robert G Endres and Ned S Wingreen. Accuracy of direct gradient sensing by cell-surface receptors. *Progress in biophysics and molecular biology*, 100(1-3):33–39, 2009.

[36] Uri Alon. *An introduction to systems biology: design principles of biological circuits*. CRC press, 2019.

[37] Brian A Camley, Yunsong Zhang, Yanxiang Zhao, Bo Li, Eshel Ben-Jacob, Herbert Levine, and Wouter-Jan Rappel. Polarity mechanisms such as contact inhibition of locomotion regulate persistent rotational motion of mammalian cells on micropatterns. *Proceedings of the National Academy of Sciences*, 111(41):14770–14775, 2014.

[38] Brian Stramer and Roberto Mayor. Mechanisms and in vivo functions of contact inhibition of locomotion. *Nature Reviews Molecular Cell Biology*, 18(1):43–55, 2017.

[39] Alex Sigal, Ron Milo, Ariel Cohen, Naama Geva-Zatorsky, Yael Klein, Yuvalal Liron, Nitzan Rosenfeld, Tamar Danon, Natalie Perzov, and Uri Alon. Variability and memory of protein levels in human cells. *Nature*, 444(7119):643–646, 2006.

[40] An alternate approach would be to consider the maximum likelihood estimation of measured signals M_i that have been time-averaged over a time τ_p , in which case h_i would be replaced in the MLE results by h_{T_i} . See Appendix D for details.

[41] Roman Vetter and Dagmar Iber. Precision of morphogen gradients in neural tube development. *Nature Communications*, 13(1):1–14, 2022.

[42] Huy Tran, Jonathan Desponts, Carmina Angelica Perez Romero, Mathieu Coppey, Cecile Fradin, Nathalie Dostatni, and Aleksandra M Walczak. Precision in a rush: Trade-offs between reproducibility and steepness of the hunchback expression pattern. *PLoS Computational Biology*, 14(10):e1006513, 2018.

[43] Carlos Carmona-Fontaine, Eric Theveneau, Apostolia Tzekou, Masazumi Tada, Mae Woods, Karen M Page, Maddy Parsons, John D Lambiris, and Roberto Mayor. Complement fragment C3a controls mutual cell attraction during collective cell migration. *Developmental Cell*, 21(6):1026–1037, 2011.

[44] Brian A Camley, Juliane Zimmermann, Herbert Levine, and Wouter-Jan Rappel. Collective signal processing in cluster chemotaxis: Roles of adaptation, amplification, and co-attraction in collective guidance. *PLoS computational biology*, 12(7):e1005008, 2016.

[45] Haicen Yue, Brian A Camley, and Wouter-Jan Rappel. Minimal network topologies for signal processing during collective cell chemotaxis. *Biophysical Journal*, 114(12):2986–2999, 2018.

[46] Bradford E Peercy and Michelle Starz-Gaiano. Clustered cell migration: Modeling the model system of drosophila border cells. In *Seminars in cell & developmental biology*, volume 100, pages 167–176. Elsevier, 2020.

[47] Austin Hopkins and Brian A Camley. Leader cells in collective chemotaxis: Optimality and trade-offs. *Physical Review E*, 100(3):032417, 2019.

[48] Rebecca McLennan, Linus J Schumacher, Jason A Morrison, Jessica M Teddy, Dennis A Ridenour, Andrew C Box, Craig L Semerad, Hua Li, William McDowell, David Kay, et al. Neural crest migration is driven by a few trailblazer cells with a unique molecular signature narrowly confined to the invasive front. *Development*, 142(11):2014–2025, 2015.

[49] Yuan Xiong, Chuan-Hsiang Huang, Pablo A Iglesias, and Peter N Devreotes. Cells navigate with a local-excitation, global-inhibition-biased excitable network. *Proceedings of the National Academy of Sciences*, 107(40):17079–17086, 2010.

[50] Andrew Mugler, Andre Levchenko, and Ilya Nemenman. Limits to the precision of gradient sensing with spatial communication and temporal integration. *Proceedings of the National Academy of Sciences*, 113(6):E689–E695, 2016.

[51] Tyler Smith, Sean Fancher, Andre Levchenko, Ilya Nemenman, and Andrew Mugler. Role of spatial averaging in multicellular gradient sensing. *Physical Biology*, 13(3):035004, 2016.

[52] Julien Varennes, Bumsoo Han, and Andrew Mugler. Collective chemotaxis through noisy multicellular gradient sensing. *Biophysical Journal*, 111(3):640–649, 2016.

[53] Danfeng Cai, Wei Dai, Mohit Prasad, Junjie Luo, Nir S Gov, and Denise J Montell. Modeling and analysis of collective cell migration in an in vivo three-dimensional environment. *Proceedings of the National Academy of Sciences*, 113(15):E2134–E2141, 2016.

[54] Loling Song, Sharvari M Nadkarni, Hendrik U Bödeker, Carsten Beta, Albert Bae, Carl Franck, Wouter-Jan Rappel, William F Loomis, and Eberhard Bodenschatz. Dictyostelium discoideum chemotaxis: threshold for directed motion. *European Journal of Cell Biology*, 85(9-10):981–989, 2006.

[55] Danny Fuller, Wen Chen, Micha Adler, Alex Groisman, Herbert Levine, Wouter-Jan Rappel, and William F Loomis. External and internal constraints on eukaryotic chemotaxis. *Proceedings of the National Academy of Sciences*, 107(21):9656–9659, 2010.

[56] Wayne A Fuller. *Measurement error models*, volume 305. John Wiley & Sons, 2009.

[57] Joseph Hesselgesser, Meina Liang, James Hoxie, Michael Greenberg, Lawrence F Brass, Michael J Orsini, Dennis Taub, and Richard Horuk. Identification and characterization of the CXCR4 chemokine receptor in human T cell lines: ligand binding, biological activity, and HIV-1 infectivity. *The Journal of Immunology*, 160(2):877–883, 1998.

[58] Jennifer L Macdonald and Linda J Pike. Heterogeneity in EGF-binding affinities arises from negative cooperativity in an aggregating system. *Proceedings of the National Academy of Sciences*, 105(1):112–117, 2008.

[59] Mario Niepel, Sabrina L Spencer, and Peter K Sorger. Non-genetic cell-to-cell variability and the consequences for pharmacology. *Current Opinion in Chemical Biology*, 13(5-6):556–561, 2009.